

# Maximising the number of induced cycles in a graph<sup>\*</sup>

Natasha Morrison<sup>†</sup>      Alex Scott<sup>†</sup>

March 15, 2016

## Abstract

We determine the maximum number of induced cycles that can be contained in a graph on  $n \geq n_0$  vertices, and show that there is a unique graph that achieves this maximum. This answers a question of Tuza [14]. We also determine the maximum number of odd or even cycles that can be contained in a graph on  $n \geq n_0$  vertices and characterise the extremal graphs. This proves a conjecture of Chvátal and Tuza from 1988 [14].

## 1 Introduction

What is the maximum number of induced cycles in a graph on  $n$  vertices? For cycles of *fixed* length, this problem has been extensively studied. Indeed, for any fixed graph  $H$ , let the *induced density* of  $H$  in a graph  $G$  be the number of induced copies of  $H$  in  $G$  divided by  $\binom{|G|}{|H|}$ ; let  $I(H; n)$  be the maximum induced density of  $H$  over all graphs  $G$  on  $n$  vertices; and let the *inducibility* of  $H$  be the limit  $\lim_{n \rightarrow \infty} I(H; n)$ . In 1975, Pippinger and Golumbic [12] made the following conjecture.

**Conjecture 1.1.** [12] *For  $k \geq 5$ , the inducibility of the cycle  $C_k$  is  $k!/(k^k - k)$ .*

Balogh, Hu, Lidický and Pfender [2] recently proved this conjecture in the case  $k = 5$  via a flag algebra method, and showed that the maximum density was achieved by a unique graph. Apart from this case, the problem remains open (though see [3, 4, 5, 6, 7, 8, 9] for results on inducibility of other graphs).

In this paper, we consider the total number of induced cycles in a graph (without restriction on length). Tuza (see [14]) asked for the maximum possible number of induced cycles in a graph with  $n$  vertices. The problem was investigated independently by Robson [10, 13], who showed in the 1980s that a graph on  $n$  vertices has at most  $3^{(1+o(1))n/3}$  induced cycles.

---

<sup>\*</sup>Work partially done during the 2015 Barbados Workshop on Structural Graph Theory. The attendance of the first author was possible due to the generous support of Merton College Oxford.

<sup>†</sup>Mathematical Institute, University of Oxford, Woodstock Road, Oxford, OX2 6GG, United Kingdom.  
E-mail: {morrison, scott}@maths.ox.ac.uk.

Tuza also asked about the number of odd induced cycles: in 1988 (see [14]) he conjectured with Chvátal that the maximum possible number of odd induced cycles in a graph on  $n$  vertices is  $3^{n/3}$ . In this paper we resolve both problems, proving exact bounds for all sufficiently large  $n$ , and determining the extremal graphs. Thus we will determine, for sufficiently large  $n$ , the graphs with  $n$  vertices that maximize the number of induced cycles; we will also determine the graphs with the maximum number of even cycles, the maximum number of odd cycles, and the maximum number of odd holes (i.e. induced odd cycles of length at least 5).

In order to state our results, it will be helpful to have a couple of definitions. As usual, for  $G$  a graph we define the *neighbourhood* of  $x$  to be  $N_G(x) := \{y \in V(G) : xy \in E(G)\}$ . We say that a graph  $\mathcal{B} := (B_1, \dots, B_k)$  is a *cyclic braid* if there exists a partition  $B_1, \dots, B_k$  of  $V(\mathcal{B})$  such that for every  $0 \leq i \leq k-1$  and every  $x \in B_i$  we have  $B_{i-1} \cup B_{i+1} \subseteq N_{\mathcal{B}}(x) \subseteq B_{i-1} \cup B_i \cup B_{i+1}$  where indices are taken modulo  $k$ . We call the sets  $B_i$  *clusters* of  $\mathcal{B}$ ; the *length* of the cyclic braid is the number of clusters. If a cyclic braid contains no edges within its clusters, we call it an *empty cyclic braid*. If it contains every edge within its clusters, we say that it is *full*. We say a pair of clusters  $C_1$  and  $C_2$  are *adjacent* in  $G$  if  $v_1v_2 \in E(G)$  for all  $v_1 \in C_1$  and  $v_2 \in C_2$ . We say a triple of clusters  $C_1, C_2, C_3$  are *consecutive* if  $C_1$  is adjacent to  $C_2$  and  $C_2$  is adjacent to  $C_3$ .

As we will see, the structure of the extremal graph will depend on the value of  $n$  modulo 3. For  $n \geq 8$  we define an  $n$ -vertex graph  $H_n$  separately for each value of  $n$  modulo 3. Let  $k \geq 3$ . We define  $H_{3k}$  to be the empty cyclic braid of length  $k$  where every cluster has size 3. We define  $H_{3k+1}$  to be the empty cyclic braid containing  $k-1$  clusters of size 3 and one of size 4. Finally, we define  $H_{3k-1}$  to be the empty cyclic braid containing  $k-1$  clusters of size 3 and one of size 2.

Let  $m(n)$  be the maximum number of induced cycles that can be contained in a graph on  $n$  vertices. We shall prove the following.

**Theorem 1.2.** *There exists  $n_0$  such that, for all  $n \geq n_0$ ,  $H_n$  is the unique graph on  $n$  vertices containing  $m(n)$  induced cycles.*

This implies immediately that  $m(n) = \Theta(3^{n/3})$ . More precisely, we get the following.

**Corollary 1.3.** *There exists  $n_0$  such that, for all  $n \geq n_0$ , we have that*

$$m(n) = \begin{cases} 3^{n/3} + 12n & \text{for } n \equiv 0 \text{ modulo } 3 \\ 4 \cdot 3^{(n-4)/3} + 12n + 51 & \text{for } n \equiv 1 \text{ modulo } 3 \\ 2 \cdot 3^{(n-2)/3} + 12n - 36 & \text{for } n \equiv 2 \text{ modulo } 3 \end{cases}$$

A *hole* is an induced cycle of length at least 4. For  $n \geq 10$ , the graph  $H_n$  does not contain any triangles, so we get the following corollary.

**Corollary 1.4.** *There exists  $n_0$  such that, for all  $n \geq n_0$ ,  $H_n$  is the unique graph on  $n$  vertices with the maximum number of holes.*

Using similar arguments to those in the proof of Theorem 1.2 we can also prove a stability-type result.

**Theorem 1.5.** *Let  $0 < \alpha < 1$  be a fixed constant. For  $n_0$  sufficiently large, let  $F$  be a graph on  $n \geq n_0$  vertices containing at least  $\alpha \cdot m(n)$  induced cycles. Then we can change edges incident to  $O(1)$  vertices of  $F$  to create a cyclic braid with the same cluster sizes as  $H_n$ .*

A stability analogue of Theorem 1.2 easily follows.

**Corollary 1.6.** *Let  $0 < \alpha < 1$  be a fixed constant. For  $n_0$  sufficiently large, let  $F$  be a graph on  $n \geq n_0$  vertices containing at least  $\alpha \cdot m(n)$  induced cycles. Then we can create a graph isomorphic to  $H_n$  by adding and deleting  $O(n)$  edges from  $F$ .*

The corollary follows directly from Theorem 1.5 and the observation that at most  $O(n)$  edges can be added to  $H_n$  in such a way that they are contained within the clusters of  $H_n$ .

The arguments used to prove Theorem 1.2 can be adapted to give results about other sets of induced cycles, for instance induced cycles of given parity. We say that a path or cycle is *odd* if it has odd length (*even* if it has even length).

Let  $m_o(n)$  be the maximum number of induced odd cycles that can be contained in a graph on  $n$  vertices. The value of  $m_o(n)$  and the structure of the extremal graphs depends on the value of  $n$  modulo 6.

Let  $G_n$  to be the full cyclic braid on  $n$  vertices whose clusters all have size 3 except for:

- three consecutive clusters of size 2, when  $n \equiv 0$  modulo 6;
- two adjacent clusters of size 2, when  $n \equiv 1$  modulo 6;
- one cluster of size 2, when  $n \equiv 2$  modulo 6;
- one cluster of size 4, when  $n \equiv 4$  modulo 6;
- two adjacent clusters of size 4, when  $n \equiv 5$  modulo 6.

We will prove the following.

**Theorem 1.7.** *There exists  $n_0$  such that, for all  $n \geq n_0$ ,  $G_n$  is the unique  $n$ -vertex graph containing  $m_o(n)$  induced odd cycles.*

Thus,  $m_o(n) = \Theta(3^{n/3})$ . If we consider odd holes (induced odd cycles of length at least 5), we get the same bound but a larger family of extremal graphs. Let  $m_o^h(n)$  be the maximum number of odd holes that can be contained in a graph on  $n$  vertices.

Define  $\mathcal{G}_n$  to be the family of cyclic braids on  $n$  vertices whose cluster sizes are the same as the cluster sizes in  $G_n$ , but with no restrictions on which clusters must be adjacent or consecutive, or on which edges are present inside the clusters; in addition, when  $n \equiv 5$  modulo 6, we also include the cyclic braids whose clusters all have size 3 except for four clusters of size 2.

A modification of the proof of Theorem 1.7 gives the following.

**Theorem 1.8.** *There exists  $n_0$  such that, for all  $n \geq n_0$ , the family of  $n$ -vertex graphs that contain  $m_o^h(n)$  odd holes is  $\mathcal{G}_n$ .*

The same techniques can be used to prove an analogous theorem for even cycles. Let  $m_e(n)$  be the maximum number of induced even cycles that can be contained in a graph on  $n$  vertices, and define  $E_n$  to be the empty cyclic braid on  $n$  vertices whose clusters all have size 3 except for:

- one cluster of size 4, when  $n \equiv 1$  modulo 6;
- two adjacent clusters of size 4, when  $n \equiv 2$  modulo 6;
- three consecutive clusters of size 2, when  $n \equiv 3$  modulo 6;
- two adjacent clusters of size 2, when  $n \equiv 4$  modulo 6; and
- one cluster of size 2, when  $n \equiv 5$  modulo 6.

The proof of Theorem 1.7 can be adapted similarly.

**Theorem 1.9.** *There exists  $n_0$  such that, for all  $n \geq n_0$ ,  $E_n$  is the unique  $n$ -vertex graph containing  $m_e(n)$  induced even cycles.*

As in the case of  $m_o(n)$ , we have  $m_e(n) = \Theta(3^{n/3})$ .

Our paper is structured as follows. In Section 2 we prove a preliminary result determining the structure of the  $n$ -vertex graphs that maximise the number of induced paths between a particular pair of vertices. During the proof of this result we will introduce several of the key ideas needed later. We then prove our main theorem in Section 3, using the preliminary result from Section 2. In Section 4 we prove Theorem 1.5, using key results from the proof of Theorem 1.2. We use very similar techniques in Section 5 to prove Theorem 1.7. From here it is easy to determine the  $n$ -vertex graphs that contain the maximum number of odd holes. Finally, in Section 6 we conclude by discussing some open questions.

## 2 Induced paths between a pair of vertices

Let  $G$  be a graph and let  $x$  and  $y$  be distinct vertices in  $V(G)$ . We define  $p_2(G; x, y)$  to be the number of induced paths in  $G$  beginning at  $x$  and ending at  $y$ . We also define:

$$p_2(G) := \max\{p_2(G; x, y) : x, y \in V(G)\},$$

and

$$p_2(n) := \max\{p_2(G) : |V(G)| = n\}.$$

Our first goal in this section is to determine the structure of the  $n$ -vertex graphs that contain  $p_2(n)$  induced paths between some pair of vertices. We will see that these extremal graphs have a particular structure that depends on the value of  $n$  modulo 3. We will then prove analogous results for odd and even length paths.

Let  $F$  be a graph and let  $B_1, \dots, B_k$  be disjoint subsets of  $V(F)$ . We say that  $\mathcal{B} := (B_1, \dots, B_k)$  is a *braid* in  $F$  if for each  $2 \leq i \leq k-1$  and for every  $x \in B_i$  we have

$B_{i-1} \cup B_{i+1} \subseteq N_F(x) \subseteq B_{i-1} \cup B_i \cup B_{i+1}$ . We refer to the sets  $B_i$  as *clusters*. If  $i \in \{1, k\}$  we say  $B_i$  is an *end cluster*; otherwise we say  $B_i$  is a *central cluster*. If  $V(F) = \bigcup_{i=1}^k B_i$ , we say that  $F$  is a *braid*. The *length* of a braid is the number of clusters it contains. As indicated in the introduction, we say that  $\mathcal{B}$  is a *cyclic braid* if for every  $0 \leq i \leq k-1$  and every  $x \in B_i$  we have  $B_{i-1} \cup B_{i+1} \subseteq N_F(x) \subseteq B_{i-1} \cup B_i \cup B_{i+1}$  where indices are taken modulo  $k$ . If a cyclic braid contains no edges within its clusters, we call it an *empty cyclic braid*. If it contains every edge within its clusters, we call it a *full cyclic braid*.

Let  $n \geq 4$ . We define  $\mathcal{F}(n)$  to be the set of all braids  $\mathcal{B}$  with the following properties.

- $|V(\mathcal{B})| = n$ .
- $\mathcal{B}$  has end clusters of size one.
- All central clusters of  $\mathcal{B}$  have size three except:
  - either a single cluster of size 4 or two clusters of size 2, when  $n \equiv 0$  modulo 3;
  - a single cluster of size 2, when  $n \equiv 1$  modulo 3.

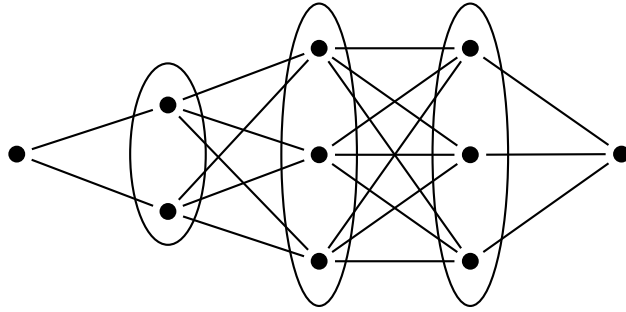


Figure 1: An example of a braid in  $\mathcal{F}(10)$ . There may or may not be edges within the clusters.

Let the end clusters be  $\{x\}$  and  $\{y\}$ . It is not difficult to check that for all  $F \in \mathcal{F}(n)$  we have  $p_2(F) = p_2(F; x, y)$ . Every graph in  $\mathcal{F}(n)$  contains the same number of induced paths between  $x$  and  $y$  and so we define

$$f_2(n) = \begin{cases} 3^{(n-2)/3} & \text{for } n \equiv 2 \text{ modulo } 3 \\ 4 \cdot 3^{(n-6)/3} & \text{for } n \equiv 0 \text{ modulo } 3 \\ 2 \cdot 3^{(n-4)/3} & \text{for } n \equiv 1 \text{ modulo } 3 \end{cases}$$

and observe that  $p_2(F) = f_2(n)$  for all  $F \in \mathcal{F}(n)$ .

We will prove the following.

**Theorem 2.1.** *Let  $G$  be a finite graph on  $n \geq 4$  vertices and let  $x$  and  $y$  be distinct vertices of  $G$ . Suppose that  $G$  contains  $p_2(n)$  induced paths between  $x$  and  $y$ . Then  $G$  is isomorphic to a graph in  $\mathcal{F}(n)$  with end clusters  $\{x\}$  and  $\{y\}$ .*

In particular, this gives the following.

**Corollary 2.2.** *For all  $n \geq 4$ , we have  $p_2(n) = f_2(n)$ .*

Before proving Theorem 2.1, we introduce some preliminary notation and definitions.

**Definition 2.3.** For a vertex  $v \in V(G)$ , let  $N^i(v)$  be the set of points at distance exactly  $i$  from  $v$ . We define  $N^k[v]$  to be the set of points within distance  $k$  of  $v$  (for example  $N^3[v] = \{v\} \cup \bigcup_{i=1}^3 N^i(v)$ ), and we define  $N[v] := N^1[v] = N(v) \cup \{v\}$ . Also, for a set  $X \subseteq V(G)$ , let  $N(X) := \bigcup_{x \in X} N(x)$ , and  $N[X] := \bigcup_{x \in X} N[x]$ . Note that  $X \subseteq N[X]$ , and  $X \cap N(X)$  may or may not be empty. For a subgraph  $H \subseteq G$ , we define  $N(H) := N(V(H))$  and  $N[H] := N[V(H)]$ .

In order to prove Theorem 2.1, we require the following concept.

**Definition 2.4.** Let  $G$  be a finite graph and  $x, y \in V(G)$ , with  $y \notin N[x]$ . The  $x$ - $y$ -path tree of  $G$  is a tree  $T = T(x, y)$  together with a function  $t : V(T) \rightarrow V(G)$  defined as follows.

- $T$  is a tree with vertex set  $V(T)$  disjoint from  $V(G)$ . The tree  $T$  is rooted at  $v_0 \in V(T)$  and we define  $t(v_0) := x$ .
- We define the tree in layers  $L_0, L_1, \dots$ , such that vertices in  $L_j$  are at distance  $j$  from  $v_0$  in  $T$ . For any vertex  $v_i \in L_i$ , we let  $P_{v_i} := v_0, v_1, \dots, v_{i-1}, v_i$  be the unique path from  $v_0$  to  $v_i$  in  $T$ . The tree will have the property that the set  $t(P_{v_i})$  induces a path in  $G$ , where we write  $t(S) := \{t(x) : x \in S\}$  for any subset  $S \subseteq V(T)$  and  $t(H) := G[\{t(x) : x \in V(H)\}]$  for any subgraph  $H \subseteq T$ .
- Given a set  $P \subseteq V(T)$ , we say that  $P$  *sees* a vertex  $w \in V(G)$  if  $w \in N_G[t(P)]$ . If  $w \notin N_G[t(P)]$ , we say that  $w$  is *unseen* by  $P$ .
- Define  $L_0 := \{v_0\}$ . Given  $L_i$ , we define  $L_{i+1}$  as follows. For each vertex  $u \in L_i$ :
  - If  $t(u) = y$ , then  $u$  is a leaf of  $T$ ; it has no neighbours in  $L_{i+1}$ .
  - If  $t(u)$  has no neighbours in  $G$  that are unseen by  $V(P_u) \setminus \{u\}$ , then  $u$  is a leaf of  $T$ .
  - If  $y$  is a neighbour of  $t(u)$  in  $G$ , then  $u$  has exactly one neighbour in  $L_{i+1}$ : a vertex  $w$  such that  $t(w) = y$ .
  - Otherwise, let  $\{w_1, \dots, w_r\} \subseteq V(G)$  be the set of vertices in  $N_G(t(u))$  unseen by  $V(P_u) \setminus \{u\}$ . We define the neighbours of  $u$  in  $L_{i+1}$  to be a set of new vertices  $\{v_1, \dots, v_r\} \subseteq L_{i+1}$  and define  $t(v_i) := w_i$ . We call the vertices in  $N(u) \cap L_{i+1}$  *children* of  $u$ .

This process must terminate as  $G$  is finite.

By construction, when  $P$  is a path in  $T$  that starts at the root  $v_0$ , we have  $t(u) \neq t(v)$  for all  $u, v \in N_T[P]$ . For any leaf  $l \in T$  such that  $t(l) = y$ , we have that  $t(P_l)$  is an induced path from  $x$  to  $y$  in  $G$ . Moreover, let  $x_0, \dots, x_l$  be any induced path in  $G$ , with  $x_0 = x$  and  $x_l = y$ . Then there exists a unique path  $v_0, \dots, v_l$  in  $T$ , satisfying  $t(v_l) = x_l$  (and  $v_l$  is a leaf of  $T$ ).

**Proposition 2.5.** *Let  $P$  be a path in  $T$  starting at  $v_0$ . If  $V(P)$  sees a vertex  $w \in V(G)$ , then there exists a unique  $u \in N_T[V(P)]$  such that  $t(u) = w$ .*

*Proof.* This follows immediately from the construction of  $T$ :  $u$  is a child in  $T$  of the first vertex  $v$  in  $P$  such that  $t(v)$  is adjacent to  $w$  in  $G$ .  $\square$

The following terminology will also be used later in the proof of Theorem 1.2.

**Definition 2.6.** Let  $T$  be an  $x$ - $y$ -path tree. Given a vertex  $u \in L_i$ , and a vertex  $w \in L_j$  for  $j > i$ , if the unique path from  $u$  to  $w$  does not contain any vertex in  $\bigcup_{k < i} L_k$ , then we say  $w$  is a *descendant* of  $u$ . For a vertex  $z \in T$ , define  $B(z)$ , the *branch rooted at  $z$* , to be the subtree induced by  $z$  and its descendants, define  $L(z)$  to be the number of leaves of  $T$  that are contained in  $B(z)$  and  $L_y(z)$  to be the number of leaves  $l$  in  $T$  contained in  $B(z)$  such that  $t(l) = y$ . Define  $C(z)$  to be the set of children of  $z$  and let  $D(z) := |C(z)|$ . If it is unclear which tree we are considering, we will write  $B_T(z)$ ,  $L_T(z)$ , etc.

In order to prove Theorem 2.1, we require the following lemma about  $x$ - $y$ -path trees.

**Lemma 2.7.** *Let  $G$  be a graph on  $n \geq 4$  vertices. Let  $x$  and  $y$  be distinct vertices in  $V(G)$ , with  $y \notin N[x]$ . Let  $T$  be the  $x$ - $y$ -path tree rooted at  $v_0$  and  $P := x_0, \dots, x_k$  be any path in  $T$  where  $x_0 = v_0$  and  $v_k$  is a leaf. Then:*

- (i)  $L_y(v_0) \leq f_2(n)$ .
- (ii) If  $L_y(v_0) = f_2(n)$ , then for any  $x_j$  and for all  $u, w \in C(x_j)$ , we have  $L_y(u) = L_y(w)$ . Also,  $t(x_k) = y$  and  $V(P) \setminus \{x_k\}$  sees every vertex of  $G$ .

*Proof.* We sequentially choose a path  $v_0, v_1, \dots, v_k \subseteq V(T)$ , where  $v_i \in L_i$  and  $v_k$  is a leaf. At vertex  $v_j$  we choose  $v_{j+1}$  to be some  $z \in C(v_j)$  such that  $L_y(z) = \max\{L_y(x) : x \in C(v_j)\}$ . Let  $\mathcal{P}$  be the set of paths that can be obtained in this manner and fix  $P := v_0, \dots, v_k \in \mathcal{P}$ . Observe that  $D(v_j) = |C(v_j)|$  is in fact the number of neighbours of  $t(v_j)$  in  $G$  unseen by  $P_{v_{j-1}}$ . For  $0 \leq i \leq k-1$ , we have

$$L_y(v_i) = \sum_{z \in C(v_i)} L_y(z) \leq D(v_i) \max\{L_y(z) : z \in C(v_i)\} = D(v_i) L_y(v_{i+1}). \quad (2.1)$$

By repeatedly applying (2.1) we get

$$L_y(v_0) \leq D(v_0) \max\{L_y(z) : z \in C(v_0)\} \leq \dots \leq \prod_{i=0}^{k-1} D(v_i). \quad (2.2)$$

By construction of  $T$ , any vertex  $u$  that has a child  $l$  with  $t(l) = y$  satisfies  $D(u) = 1$ . Thus

$$L_y(v_0) \leq \prod_{i=0}^{k-2} D(v_i), \quad (2.3)$$

where  $\sum_{i=0}^{k-2} D(v_i) \leq n - 2$ , as  $v_0, \dots, v_{k-2}$  have disjoint sets of children in  $G \setminus \{x, y\}$ .

It is easily checked that the maximal value of  $\prod_{i=0}^{k-2} D(v_i)$  subject to  $\sum_{i=0}^{k-2} D(v_i) \leq n - 2$  occurs only in the following cases:

- If  $n \equiv 2$  modulo 3, we have  $D(v_i) = 3$  for all  $i$ .
- If  $n \equiv 0$  modulo 3, we have either  $D(v_i) = 4$  for exactly one  $i$  and  $D(v_j) = 3$  for all  $j \neq i$ ; or there are  $i_1, i_2$  such that  $D(v_i) = 2$  for  $i = i_1, i_2$ , and  $D(v_j) = 3$  for all  $i \notin \{i_1, i_2\}$ .
- If  $n \equiv 1$  modulo 3, we have  $D(v_i) = 2$  for exactly one  $i$ , and  $D(v_j) = 3$  for all  $j \neq i$ .

Thus we see that the maximal possible value of  $\prod_{i=0}^{k-2} D(v_i)$  is  $f_2(n)$ , and so  $L_y(v_0) \leq f_2(n)$  as required for (i).

When  $L_y(v_0) = f_2(n)$  we have that  $\prod_{i=0}^{k-2} D(v_i)$  must equal  $f_2(n)$ . This is only possible if  $\sum_{i=0}^{k-2} D(v_i) = n - 2$  and the  $D(v_i)$  take the values defined in the above cases. In addition, we must have equality in (2.3) and hence in (2.1) for each value of  $0 \leq i \leq k - 1$ . Therefore, for each  $i$  and for all  $z, w \in C(v_i)$ , we have  $L_y(z) = L_y(w)$ .

Let  $X := x_0, \dots, x_k$  be any path where  $x_0 = v_0$  and  $x_k$  is a leaf. We prove that

$$\prod_{i=0}^{k-2} D(x_i) = f_2(n). \quad (2.4)$$

Choose  $P' := y_0, \dots, y_k \in \mathcal{P}$  so that it coincides with  $X$  on the longest possible initial segment, i.e. so that  $i$  is maximal such that  $y_0, \dots, y_i = x_0, \dots, x_i$ . Suppose  $P' \neq X$ . Then for some  $j$  we have  $L_y(x_j) \neq L_y(y_j)$ , but  $x_i = y_i$  for  $i < j$ . But by the argument above, as  $P \in \mathcal{P}$ , we have that for each  $i$ ,  $L_y(z) = L_y(w)$  for all  $z, w \in C(y_i)$ . Thus as  $x_{j-1} = y_{j-1}$ , we have  $x_j \in C(y_{j-1})$  and  $L_y(x_j) = L_y(y_j)$ , a contradiction. So  $P' = X$  and every path terminating at a leaf is in  $\mathcal{P}$ . So for any  $x_j$  and any  $u, w \in C(x_j)$  we have  $L_y(u) = L_y(w)$ .

By choice of  $P'$  we have  $\prod_{i=0}^{k-2} D(y_i) = \prod_{i=0}^{k-2} D(v_i) = f_2(n)$ . Hence (2.4) holds for  $X$ . But then we must have  $\sum_{j=0}^{k-2} D(x_j) = n - 2$  and so  $X \setminus \{x_k\}$  must see every vertex of  $G$  as required. But then  $t(x) = y$  for some  $x \in N_T[X]$ , and the construction of  $T$  implies that we must have  $x = x_k$  (as  $x$  must be a leaf and the parent of  $x$  has only one child).  $\square$

We are now ready to prove Theorem 2.1.

*Proof of Theorem 2.1.* Let  $T := T(x, y)$  be the  $x$ - $y$ -path tree of  $G$  rooted at  $v_0$ . The number of induced paths between  $x$  and  $y$  is precisely the number of leaves  $l \in T$  such that  $t(l) = y$ . So  $L_y(v_0) = f_2(n)$  and we can apply Lemma 2.7. We intend to show that  $G$  must be a



braid. Let  $P := v_0, \dots, v_k$  be the shortest path in  $T$  such that  $t(v_0) = x$  and  $t(v_k) = y$ . For  $i \in \{0, \dots, k-1\}$ , let  $C_{i+1}$  be the set of children of  $v_i$  in  $T$  (note that  $v_{i+1} \in C_{i+1}$ ). Define  $V_0 := \{x\}$  and define  $V_i := t(C_i)$  for  $1 \leq i \leq k$ . Therefore,  $V_k = \{y\}$ . The sets  $V_i$  are disjoint, as  $t(u) \neq t(u')$  for any  $u, u' \in N_T(\{v_0, \dots, v_k\})$ . We also have that  $\bigcup_{i=0}^k V_i = V(G)$ , as  $V(P)$  sees every vertex in  $G$  by Lemma 2.7. The theorem will follow immediately from the next claim.

**Claim 2.8.**  $G$  is the braid  $(\{x\}, V_1, \dots, V_{k-1}, \{y\})$ .

*Proof.* We prove by reverse induction on  $j$  that the graph induced by  $\bigcup_{i=j}^k V_i$  is a braid  $(V_j, \dots, V_k)$  in  $G$ . First, note that by Lemma 2.7, every leaf  $l \in T$  satisfies  $t(l) = y$ . Thus no vertex in  $C_{k-1}$  can be a leaf of  $T$  (else we would have a shorter path to  $y$  in  $G$ ) and every vertex in  $C_{k-1}$  must have a child. Since, by Lemma 2.7,  $V(P) \setminus \{v_k\}$  sees every vertex of  $G$  and all vertices except  $y$  have been seen by  $v_0, \dots, v_{k-2}$ , every child  $z$  of a vertex in  $C_{k-1}$  must satisfy  $t(z) = y$ . Thus every vertex in  $C_{k-1}$  has exactly one child  $z$  and  $t(z) = y$ . Therefore  $(V_{k-1}, \{y\})$  is a braid.

Now suppose that the inductive hypothesis holds for  $j = s+1$  where  $1 \leq s+1 \leq k-1$ . So  $(V_{s+1}, \dots, V_k)$  is a braid in  $G$ . We will show that  $(V_s, \dots, V_k)$  is a braid in  $G$ . We first show there are no edges between  $\bigcup_{i=0}^{s-1} V_i$  and  $V_{s+1}$ . Suppose, for some  $i \leq s-1$ , there exists a vertex  $v \in L_i$  with a child  $z$  such that  $t(z) \in V_{s+1}$ . Then there exists a shorter path in  $T$  from  $v_0$  to some  $v_k$  where  $t(v_k) = y$ : the path  $v_0, \dots, v_{i-1}, v, z, v_{s+2}, \dots, v_k$ . This contradicts our choice of  $P$  as the shortest such path. Thus no such vertex  $v$  exists. So there are no edges between  $\bigcup_{i=0}^{s-1} V_i$  and  $V_{s+1}$ .

It remains to show that  $\{uw : u \in V_s, w \in V_{s+1}\} \subseteq E(G)$ . Suppose there exists some  $v \in C_s \setminus \{v_s\}$  and some  $z \in C_{s+1}$  such that  $t(v)$  is not adjacent to  $t(z) \in V_{s+1}$ .

Let  $u$  be a child of  $v$  ( $v$  must have a child as, by Lemma 2.7, every leaf  $l$  satisfies  $t(l) = y$ ). We know that  $t(u)$  is a neighbour of  $t(v)$  in  $G$  and that:

- $t(u) \notin \bigcup_{i=0}^s V_i$ , as  $t(u)$  must be unseen by  $\{v_0, \dots, v_{s-1}\}$  by construction of  $T$ ;
- $t(u) \notin \bigcup_{i=s+2}^k V_i$ , as  $V_{s+1}, \dots, V_k$  forms a braid in  $G$ .

Thus  $t(u) \in V_{s+1}$ .

If  $s+2 = k$  (and so  $V_{s+2} = V_k = \{y\}$ ) then consider the path  $v_0, v_1, \dots, v_{s-1}, v \in T$ . We have  $L_y(v) = D(v) < D(v_s) = L_y(v_s)$ , contradicting Lemma 2.7.

Therefore  $s+2 < k$ . Since every leaf  $l \in T$  satisfies  $t(l) = y$ , every induced path starting at  $x$  in  $G$  can be extended to an induced path terminating at  $y$ . We consider two cases (see Figure 2 for an illustration).

First suppose that  $t(u)$  is adjacent to  $t(z)$ . Consider  $P := t(v_0), \dots, t(v_{s-1}), t(v), t(u), t(z)$ , an induced path in  $G$ . As  $t(z) \in V_{s+1}$  and  $(V_{s+1}, \dots, V_k)$  is a braid in  $G$ , any extension of  $P$  to an induced path that terminates at  $y$  must contain a vertex from  $V_{s+2}$ . However,  $V_{s+2} \subseteq N_G(t(u))$  (and so  $V_{s+2}$  has been seen by  $V(P_u)$ ). It is therefore impossible to extend this induced path to one terminating at  $y$ .

Now suppose that  $t(u)$  is not adjacent to  $t(z)$ . Let  $w$  be a neighbour of  $t(u)$  in  $V_{s+2}$ . Consider the induced path  $P := t(v_0), \dots, t(v_{s-1}), t(v), t(u), w, t(z)$ . As  $t(z) \in V_{s+1}$  and

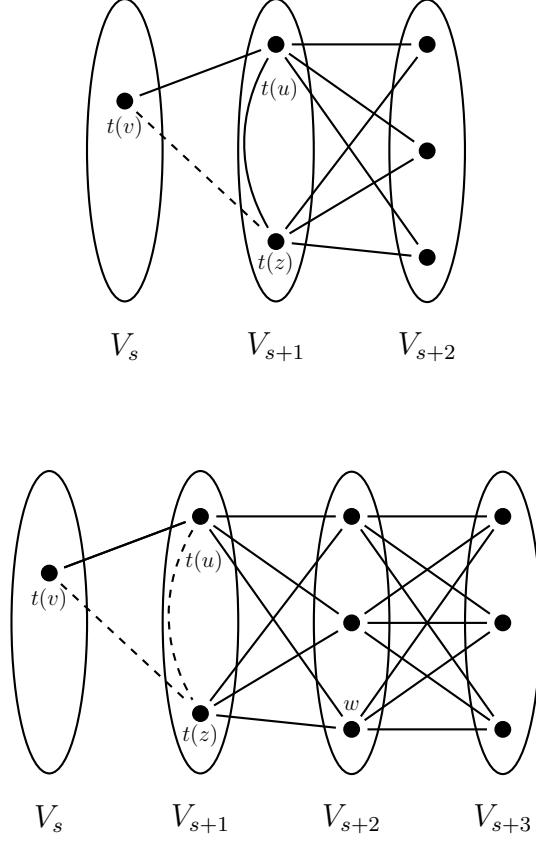


Figure 2: Examples of the cases we get if  $t(v)$  is not adjacent to every vertex in  $V_{s+1}$ . The dashed lines represent non-edges.

$(V_{s+1}, \dots, V_k)$  is a braid in  $G$ , any extension of  $P$  to an induced path that terminates at  $y$  must contain a vertex from  $V_{s+3}$ . However,  $V_{s+3} \subseteq N_G(w)$  and so has been seen by  $V(P_w)$ . It is therefore impossible to extend this induced path to one terminating at  $y$ .

Thus it must be the case that  $\{uw : u \in V_s, w \in V_{s+1}\} \subseteq E(G)$ . We conclude that the graph induced by  $\bigcup_{i=s}^k V_i$  is indeed a braid  $(V_s, \dots, V_k)$  in  $G$ . Claim 2.8 now follows by induction.  $\square$

We have  $|V_i| = D(v_{i-1})$ . As  $\prod_{i=0}^{k-2} D(v_i) = p_2(F(n))$ , it is easy to check that the braid must be in  $\mathcal{F}(n)$ . Hence the theorem follows.  $\square$

## 2.1 Odd and even induced paths between a pair of vertices

Let  $G$  be a graph and let  $x$  and  $y$  be distinct vertices in  $V(G)$ . We will define similar notions for odd and even paths as we did for paths in general at the start of this section. We define  $p_2^o(G; x, y)$  to be the number of induced odd paths in  $G$  beginning at  $x$  and ending at  $y$ . We also define:

$$p_2^o(G) := \max\{p_2^o(G; x, y) : x, y \in V(G)\},$$

and

$$p_2^o(n) := \max\{p_2^o(G) : |V(G)| = n\}.$$

In addition, we define  $p_2^e(G; x, y)$  to be the number of induced even paths in  $G$  beginning at  $x$  and ending at  $y$ . In the even case, we define  $p_2^e(G)$  and  $p_2^e(n)$  analogously to  $p_2^o(G)$  and  $p_2^o(n)$ .

We will determine the structure of the  $n$ -vertex graphs that contain  $p_2^o(n)$  induced odd paths (or  $p_2^e(n)$  induced even paths) between some pair of vertices. This result will be used to prove Theorem 1.7. The extremal graphs for this path problem will have a certain structure that depends on the value of  $n$  modulo 6. For  $n \geq 10$ , we define  $\mathcal{F}_o(n)$  to be the set of all braids  $\mathcal{B}$  with the following properties.

- $|V(\mathcal{B})| = n$ .
- $\mathcal{B}$  has end clusters of size 1.
- All central clusters of  $\mathcal{B}$  have size three except:
  - a single cluster of size 4, when  $n \equiv 0$  modulo 6;
  - either two clusters of size 4 or four clusters of size 2, when  $n \equiv 1$  modulo 6;
  - three clusters of size 2, when  $n \equiv 2$  modulo 6;
  - two clusters of size 2, when  $n \equiv 3$  modulo 6; and
  - one cluster of size 2, when  $n \equiv 4$  modulo 6.

Let  $F \in \mathcal{F}_o(n)$  and suppose that the end clusters are  $\{x\}$  and  $\{y\}$ . Observe that every induced path between  $x$  and  $y$  is odd. It is not difficult to check that for all  $F \in \mathcal{F}_o(n)$  we have  $p_2^o(F) = p_2^o(F; x, y)$ . Every graph in  $\mathcal{F}_o(n)$  contains the same number of induced paths between  $x$  and  $y$  and so we define:

$$f_2^o(n) = \begin{cases} 4 \cdot 3^{(n-6)/3} & \text{for } n \equiv 0 \text{ modulo } 6 \\ 2^4 \cdot 3^{(n-10)/3} & \text{for } n \equiv 1 \text{ modulo } 6 \\ 2^3 \cdot 3^{(n-8)/3} & \text{for } n \equiv 2 \text{ modulo } 6 \\ 2^2 \cdot 3^{(n-6)/3} & \text{for } n \equiv 3 \text{ modulo } 6 \\ 2 \cdot 3^{(n-4)/3} & \text{for } n \equiv 4 \text{ modulo } 6 \\ 3^{(n-2)/3} & \text{for } n \equiv 5 \text{ modulo } 6 \end{cases}$$

The following theorem is the analogous theorem for odd paths to Theorem 2.1. The proof is very similar.

**Theorem 2.9.** *Let  $G$  be a finite graph on  $n \geq 10$  vertices and let  $x$  and  $y$  be distinct vertices of  $G$ . Suppose that  $G$  contains  $p_2^o(n)$  induced odd paths between  $x$  and  $y$ . Then  $G$  is isomorphic to a graph in  $\mathcal{F}_o(n)$  with end clusters  $\{x\}$  and  $\{y\}$ .*

We will state a version of Theorem 2.9 for even length paths. As one would expect, the extremal graphs differ from those in the odd case. Thus for  $n \geq 10$ , we define  $\mathcal{F}_e(n)$  to be the set of all braids  $\mathcal{B}$  with the following properties.

- $|V(\mathcal{B})| = n$ .
- $\mathcal{B}$  has end clusters of size 1.
- All central clusters of  $\mathcal{B}$  have size three except:
  - two clusters of size 2, when  $n \equiv 0$  modulo 6; and
  - one cluster of size 2, when  $n \equiv 1$  modulo 6.
  - a single cluster of size 4, when  $n \equiv 3$  modulo 6;
  - either two clusters of size 4 or four clusters of size 2, when  $n \equiv 4$  modulo 6; and
  - three clusters of size 2, when  $n \equiv 5$  modulo 6.

Observe that the extremal graphs in the odd and even cases are essentially the same (shifting by 3 modulo 6), as when  $n \geq 13$  we can delete a cluster of size 3 to get from an extremal graph for the odd case to an extremal graph for the even case (or vice versa).

**Theorem 2.10.** *Let  $G$  be a finite graph on  $n \geq 10$  vertices and let  $x$  and  $y$  be distinct vertices of  $G$ . Suppose that  $G$  contains  $p_2^e(n)$  induced even paths between  $x$  and  $y$ . Then  $G$  is isomorphic to a graph in  $\mathcal{F}_e(n)$  with end clusters  $\{x\}$  and  $\{y\}$ .*

To obtain the proof of Theorem 1.7, it suffices to prove only the odd version. We remark that the proof of Theorem 2.9 can easily be adapted to prove Theorem 2.10, so we omit the proof of Theorem 2.10.

*Sketch proof of Theorem 2.9.* Fix  $x$  and  $y$  to be distinct vertices of  $G$  with  $y \notin N[x]$ . Let  $T$  be the  $x$ - $y$ -path tree rooted at  $v_0$ . For  $z \in V(T)$  define  $L_o(z)$  to be the number of leaves  $l$  contained in  $B(z)$  such that  $t(l) = y$  and  $l \in L_r$ , where  $r$  is even (as we start at level 0, a path terminating at such a level will be odd).

We first prove an *odd-path version* of Lemma 2.7.

**Claim 2.11** (Odd-path version of Lemma 2.7). Let  $P := x_1, \dots, x_k$  be any path in  $T$  where  $x_0 = v_0$  and  $x_k$  is a leaf. Then:

- (i)  $L_o(v_0) \leq f_2^o(n)$ .
- (ii) If  $L_o(v_0) = f_2^o(n)$ , then for any  $x_j$  and for all  $u, w \in C(x_j)$ , we have  $L_o(u) = L_o(w)$ . Also,  $k$  is even,  $t(x_k) = y$  and  $V(P) \setminus \{x_k\}$  sees every vertex of  $G$ .

*Proof of Claim.* We mimic the proof of Lemma 2.7, replacing  $\mathcal{F}(n)$  with  $\mathcal{F}_o(n)$  and  $L_y(z)$  with  $L_o(z)$  for any  $z \in T$ .

Arguing as in (2.2), we see that

$$L_o(v_0) \leq D(v_{k-1}) \cdot \prod_{i=0}^{k-2} D(v_i). \quad (2.5)$$

By construction of  $T$ , any vertex  $u$  that has a child  $l$  with  $t(l) = y$  satisfies  $D(u) = 1$ . If  $k - 2$  is odd, any leaf in  $L_k$  corresponds to an even path, so  $k$  must be even. So we have

$$L_o(v_0) \leq \prod_{i=0}^{k-2} D(v_i),$$

where  $k$  is even and  $\sum_{i=0}^{k-2} D(v_i) \leq n - 2$ , as  $v_0, \dots, v_{k-2}$  have disjoint sets of children in  $G \setminus \{x, y\}$ .

It is not difficult to check that the maximal value of  $\prod_{i=0}^{k-2} D(v_i)$  subject to  $\sum_{i=0}^{k-2} D(v_i) \leq n - 2$ , where  $k$  is even is  $f_2^o(n)$ .

The proof of the second statement in Claim 2.11 follows directly from the arguments used in the proof of the second statement of Lemma 2.7. This completes the proof of the claim.  $\square$

In particular, we know that any path  $v_0 \dots v_k$ , where  $v_k$  is a leaf, has odd length and satisfies  $t(v_k) = y$ . We can now use the same arguments that we used in the proof of Theorem 2.1, applying Claim 2.11 in the place of Lemma 2.7, to show that  $G$  must be a braid in  $\mathcal{F}_o(n)$ .  $\square$

### 3 Proof of Theorem 1.2

We fix a large constant  $n_0$  and let  $G_{max}$  be a graph on  $n \geq n_0$  vertices, that contains  $m(n)$  induced cycles. In what follows we will take  $n_0$  (and thus  $n$ ) to be sufficiently large when required and we will make no attempts to optimise the constants in our arguments. We will show that the graph  $G_{max}$  is isomorphic to  $H_n$ . As it turns out, Theorem 1.5 will follow almost immediately from the arguments required for the proof of Theorem 1.2. Therefore, in this section we will prove several lemmas in more generality than is strictly needed here: they will be used in their more general form in the next section.

Given a graph  $H$ , let  $f(H)$  denote the number of induced cycles in  $H$  and for a vertex  $v \in H$ , let  $f_v(H)$  denote the number of induced cycles in  $H$  containing  $v$ . Observe that we must have:

$$f(G_{max}) = m(n) \geq f(H_n) \geq \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \text{ modulo } 3 \\ 4 \cdot 3^{(n-4)/3} & \text{if } n \equiv 1 \text{ modulo } 3 \\ 2 \cdot 3^{(n-2)/3} & \text{if } n \equiv 2 \text{ modulo } 3. \end{cases} \quad (3.1)$$

It is easy to see that  $G_{max}$  must be connected. We begin by proving several lemmas which determine information about the structure of  $G_{max}$ .

**Lemma 3.1.** *Let  $F$  be an  $n$ -vertex graph. For  $v \in V(F)$ , we have  $f_v(F) \leq \binom{d(v)}{2} 3^{(n-d(v)-1)/3}$ .*

*Proof.* Each induced cycle containing  $v$  contains exactly two neighbours of  $v$ . Fix a pair of vertices  $u, w \in N(v)$ . By Corollary 2.2 there are at most  $3^{(n-d(v)-1)/3}$  induced paths between  $u$  and  $w$  in  $(F \setminus N[v]) \cup \{u, w\}$ . Thus there can be at most  $3^{(n-d(v)-1)/3}$  induced cycles in  $F$  containing  $\{v, u, w\}$ . As there are  $\binom{d(v)}{2}$  distinct pairs of neighbours of  $v$ , we have,

$$f_v(F) \leq \binom{d(v)}{2} 3^{(n-d(v)-1)/3},$$

as required.  $\square$

**Lemma 3.2.** *Let  $0 < c \leq 1$ . Let  $F$  be an  $n$ -vertex graph with  $f(F) \geq c \cdot m(n)$ . Then  $F$  contains a vertex  $v$  such that  $f_v(F) \geq \frac{c}{10}m(n)$ . Moreover, every vertex  $w \in G_{max}$  satisfies  $f_w(G_{max}) \geq \frac{1}{20}m(n)$ .*

*Proof.* First we show that for large  $n$  almost all cycles in  $F$  have length at least  $0.11n$ . Let  $\alpha = 0.11$ . Using Stirling's approximation, we get that

$$\sum_{i=1}^{\lfloor \alpha n \rfloor} \binom{n}{i} \leq \alpha n \cdot \binom{n}{\alpha n} \leq (1 + o(1)) \frac{\sqrt{\alpha n}}{\sqrt{2\pi(1-\alpha)}} \left[ \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} \right]^n.$$

As

$$\frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} < 3^{1/3},$$

we get that

$$\sum_{i=1}^{\lfloor \alpha n \rfloor} \binom{n}{i} = o(3^{n/3}).$$

Thus, as  $F$  has at least  $\Omega(3^{n/3})$  induced cycles by (3.1), for sufficiently large  $n_0$ , at least  $0.99c \cdot m(n)$  cycles have length at least  $\alpha n$  as required. In particular, provided  $n_0$  is sufficiently large, we have

$$\sum_{i=1}^{\lfloor \alpha n \rfloor} \binom{n}{i} < \frac{c}{100} \cdot 3^{n/3} < \frac{c}{100} m(n),$$

for all  $n > n_0$ . The second inequality follows from (3.1). Thus there exists a vertex  $v$  such that

$$f_v(F) \geq \frac{99\alpha c}{100} m(n) \geq \frac{c}{10} m(n).$$

Now suppose that there exists some vertex  $w \in G_{max}$  with  $f_w(G_{max}) < \frac{1}{20}m(n)$ . Consider the graph  $G'$  obtained from  $G_{max}$  by duplicating the vertex  $v$  and removing the vertex  $w$ . We have that

$$f(G') \geq f(G_{max}) + f_v(G_{max}) - 2f_w(G_{max}) > f(G_{max}),$$

a contradiction.  $\square$

Now consider the graph obtained from  $G_{max}$  by duplicating a vertex. By applying the result of Lemma 3.2, it is easy to see that

$$m(n+1) \geq \left(1 + \frac{1}{20}\right) m(n). \quad (3.2)$$

We now use this to show that the graph  $G_{max}$  has maximum degree bounded by a constant.

**Lemma 3.3.**  $\Delta(G_{max}) < 30$ .

*Proof.* Let  $v$  be a vertex of maximal degree. Given  $v$ , we can split the induced cycles in  $G_{max}$  into those contained in  $G_{max} \setminus \{v\}$ , and those containing  $v$ . Using Lemma 3.1 we have

$$m(n) \leq m(n-1) + \binom{d(v)}{2} 3^{(n-d(v)-1)/3}.$$

Using (3.2) to bound  $m(n-1)$  gives,

$$m(n) \leq m(n) \left(1 + \frac{1}{20}\right)^{-1} + \binom{d(v)}{2} 3^{(n-d(v)-1)/3}.$$

This expression rearranges to give

$$m(n) \leq 21 \binom{d(v)}{2} 3^{(n-d(v)-1)/3}.$$

For  $d(v) \geq 30$ , we get that  $m(n) < 3^{(n-6)/3}$ , a contradiction (as we have  $m(n) \geq f(H_n) > 3^{(n-6)/3}$ ).  $\square$

Combining Lemma 3.1 with Lemma 3.3 shows, for any  $v \in V(G_{max})$ , we have  $f_v(G_{max}) \leq \binom{30}{2} 3^{(n-3)/3}$ . This along with (3.1) implies

$$m(n) = \Theta(3^{n/3}). \quad (3.3)$$

The next stage of our proof involves showing that all but a constant number of vertices  $v$  in our graph have the property that their closed third neighbourhood  $N^3[v]$  has the same local structure as the closed third neighbourhood of a vertex in  $H_n$ . We introduce some preliminary definitions.

Given a graph  $F$  and a set  $S \subseteq V(F)$ , we say that a vertex  $v \in V(F)$  is *seen by*  $S$  if  $v \in N[S]$  and  $v$  is *unseen by*  $S$  otherwise. Given a subgraph  $H \subseteq F$ , we say  $v$  is seen by  $H$  if  $v \in N[H]$ . When it is clear which set/subgraph we are referring to, we will just say  $v$  is (un)seen. In order to determine what sort of local structure a ‘typical’ vertex in  $G_{max}$  should have, we define a game on  $F$ .

**Definition 3.4.** Let  $F$  be a finite graph, let  $v \in V(F)$  and let  $w \in V(F) \setminus N^3[v]$ . We define the  $w$ -typical-game on  $(F, v)$  as follows. There are two players, *Adversary* and *Builder*. The game starts at vertex  $u_1 := v$  and the players choose a sequence of vertices  $\{u_2, \dots, u_k\}$  under the following set of rules. At vertex  $u_i$ :

- If  $u_i \in N^3[w]$ , then Adversary is the *active player*, otherwise Builder is.
- The active player chooses a neighbour  $u_{i+1}$  of  $u_i$  that is unseen by  $\{u_1, \dots, u_{i-1}\}$ .
- The game terminates when a vertex  $u_j$  is chosen such that  $u_j$  has no neighbours unseen by  $\{u_1, \dots, u_{j-1}\}$ .
- If, for some  $j$ , the vertex  $u_j$  does not have exactly 3 neighbours unseen by  $\{u_1, \dots, u_{j-1}\}$ , we call  $u_j$  *bad*; we call  $u_j$  *good* otherwise.
- Adversary wins if either:
  - for some  $j$ , the vertex  $u_j$  is in  $N^3[w]$  and is bad; or
  - upon termination of the game at vertex  $u_k$ , there exists a vertex in  $N^3[w]$  that is unseen by  $\{u_1, \dots, u_k\}$ .

Builder wins otherwise.

We say that a vertex  $w \in V(F) \setminus N^3[v]$  is *v-typical* in  $F$  if there exists a winning strategy for Builder in the  $w$ -typical-game on  $(F, v)$ . A vertex is *v-atypical* otherwise. When it is clear which vertex has been chosen to play the role of  $v$ , we simply say that  $w$  is (a)typical. Note that the set of vertices  $\{u_1, \dots, u_k\}$  chosen during the game induces a path in  $F$ . Also, if we play this game on  $H_n$  starting at any vertex, the vertices that are chosen are mostly good.

Now we see that a  $v$ -typical vertex has the local structure we require.

**Lemma 3.5.** *Let  $F$  be a graph and let  $v$  be a vertex in  $F$ . Suppose that  $z = z_1 \in V(F) \setminus N^3[v]$  is a  $v$ -typical vertex. Then there exist disjoint sets of vertices  $Z := \{z_1, z_2, z_3\}$ ,  $V := \{v_1, v_2, v_3\}$ , and  $W := \{w_1, w_2, w_3\}$  such that for all  $i$ , we have  $V \cup W \subseteq N(z_i) \subseteq V \cup W \cup Z \setminus \{z_i\}$ . Moreover, for  $i, j \in \{1, 2, 3\}$ , we have  $N(v_i) \cap N(w_j) = Z$ .*

*Proof.* We play as Adversary in the  $z$ -typical-game on  $(F, v)$ . As  $z$  is  $v$ -typical, Builder has a winning strategy  $\sigma$ . We assume that Builder uses strategy  $\sigma$ , and deduce information about the structure of  $F$  from the results of our choices of vertices as Adversary (we know we cannot win so whatever choices we make have certain consequences). For each vertex  $u_i$  that is chosen, let  $P_{u_i}$  denote the subgraph induced by  $\{u_1, \dots, u_i\}$ , where  $u_1 = v$ . So  $P_{u_i}$  is an induced path between  $v$  and  $u_i$ .

Suppose that  $u_k$  is the first vertex chosen such that  $u_k \in N^3(z)$  (as  $z$  is typical, at some point such a vertex must be chosen). We arbitrarily choose the next vertex  $u_{k+1} \in N^2(z) \cap N(u_k)$ . Let  $x := u_{k+1}$ . This vertex is unseen by  $P_{u_{k-1}}$  as  $u_k$  was the first vertex we chose in  $N^3(z)$ . We also have, by choice of  $u_k$ , that  $x \notin N(z)$ . As  $z$  is typical,  $x$  must have 3 neighbours  $V := \{v_1, v_2, v_3\}$  unseen by  $P_{u_k}$ . As  $x \in N^2(z)$ , for some  $i$  we must have that  $v_i \in N(x) \cap N(z)$ . Without loss of generality we suppose  $v_1 \in N(x) \cap N(z)$ . Since we could choose  $u_{k+2}$  to be  $v_1$ , the vertex  $v_1$  must have 3 neighbours unseen by  $P_x$ , one of which must be  $z$ , so let the set of neighbours of  $v_1$  unseen by  $P_x$  be  $Z := \{z_1, z_2, z_3\}$ .



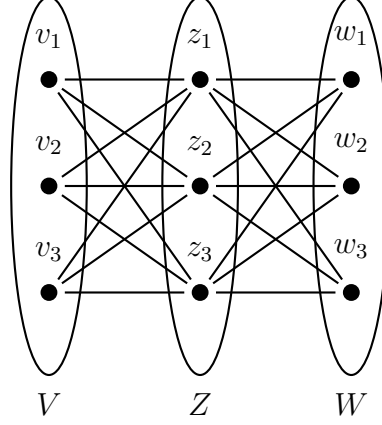


Figure 3: The local structure around a  $v$ -typical vertex  $z_1$ . Note that we have not yet determined the edges within the sets  $V$ ,  $Z$  and  $W$ .

If we choose  $u_{k+2} := v_1$  and then  $u_{k+3} := z$ , we have that  $z$  must have 3 neighbours unseen by  $P_{v_1}$ . Thus  $z$  has at most 5 neighbours unseen by  $P_x$  (as  $z$  could be adjacent to  $z_2$  or  $z_3$ ).

We now prove that  $N(x) \cap N(z) = V$ . Suppose otherwise, so without loss of generality we have  $v_3 \notin N(z)$ . Now we describe the set of choices we make for the remainder of the game (recall that Builder always plays by strategy  $\sigma$ ). Choose  $u_{k+2} := v_3$ . Now consider a later step in the game, but before  $z$  has been chosen, and suppose the most recently chosen vertex is  $u_i \in N^3[z]$ , where  $i \geq k+2$ . Then:

- (1) If there is no vertex in  $N(u_i) \cap N(z)$  that is unseen by  $P_{u_{i-1}}$ , choose  $u_{i+1}$  arbitrarily.
- (2) If, for some  $r$ , the vertex  $z$  has  $r$  neighbours unseen by  $P_{u_{i-1}}$  (by the argument above,  $r \leq 5$ ) and  $u_i$  is adjacent to  $j \geq 1$  of these neighbours:
  - (i) If  $r - j < 3$ , choose  $u_{i+1} \in N(u_i) \cap N(z)$  and  $u_{i+2} := z$ . As  $z$  now has at most 2 neighbours unseen by  $P_{u_{i+1}}$  we reach a contradiction.
  - (ii) If  $r - j \geq 3$ , then  $j \leq 2$  and  $u_i$  must have an unseen neighbour  $s$  that is not adjacent to  $z$  (otherwise  $u_i$  is bad, which contradicts the fact that Builder has a winning strategy). Choose  $u_{i+1} := s$ . Observe that the vertex  $z$  is not seen by  $P_s$ .

Once we have chosen  $z$ , we play arbitrarily.

We now analyse the results of making these choices. As  $z$  is typical, we must at some point enter case (2). If we are in case (2ii), we pick  $u_{i+1}$  and the number of neighbours unseen by  $z$  decreases, so eventually we must enter case (2i) where we reach a contradiction. Thus  $N(x) \cap N(z) = V$ , as required.

We now know that  $N(x) \cap N(z) = V$ . As in the first part, if we choose  $u_{k+2} := v_1$  and  $u_{k+3} := z$ , the vertex  $z$  must have 3 neighbours unseen by  $P_{v_1}$ . Call these neighbours

$W := \{w_1, w_2, w_3\}$ . Thus we have  $N(z) \supseteq V \cup W$  as required. From the argument above it is clear that  $N(z) \subseteq V \cup W \cup Z \setminus \{z\}$ .

We now prove the second claim of the lemma. Let  $w_j \in W$ , for  $j \in \{1, 2, 3\}$ . We will first show that  $N(v_1) \cap N(w_j) = Z$ . Again we play as Adversary in a new  $z$ -typical-game on  $(F, v)$ . We assume that Builder uses the same winning strategy  $\sigma$  as above.

By playing the same strategy as in the game above, we have that  $u_k$  is the first vertex chosen in  $N^3(z)$  and that the first vertex chosen in  $N^2(z)$  is  $x$ . We know that  $N(x) \cap N(z) = V$ . Now we choose  $u_{k+2} := v_1$ . Let  $y$  be the first vertex chosen in  $N^2(w_j)$ . Observe that as  $w_j \in N(z)$ , we have  $N^2(w_j) \subseteq N^3[z]$ . So  $y \in \{u_k, x, v_1\}$ . We will show that  $y = v_1$ .

By repeating the same argument as above with  $y$  in place of  $x$  and  $w_j$  in place of  $z$ , we see that  $y$  and  $w_j$  must have exactly 3 common neighbours. The assumptions we needed to apply the argument above will hold here: we require that  $y$  is the first vertex seen in  $N^2(w_j)$  and that we are able to make moves at vertices in  $N^2[w_j]$ . The latter holds as  $N^2[w_j] \subseteq N^3[z]$  and, in the  $z$ -typical game on  $(F, v)$ , Adversary makes moves at vertices in  $N^3[z]$ .

Suppose that  $y = u_k$ . Then  $|N(u_k) \cap N(w_j)| = 3$  and in particular  $w_j$  must be adjacent to  $x$ . This is impossible as  $w_j \notin V$ . Now suppose that  $y = x$ . In this case  $w_j$  must be adjacent to  $v_1$ . As  $w_j \notin Z$ , this is impossible. Therefore  $y = v_1$ .

It remains to show that, for  $i, j \in \{2, 3\}$ , each  $v_i$  is adjacent to each  $z_j$ . Suppose that  $v_i$  is not adjacent to  $z_j$  for some  $i, j \in \{2, 3\}$ . Again we play as Adversary in a new  $z$ -typical-game on  $(F, v)$ . We assume that Builder uses the same winning strategy  $\sigma$  as above. By playing the same strategy as in both games above, we can ensure that the first vertex chosen in  $N^2[z]$  is  $x$  and that  $x = u_{k+1}$  in the sequence of vertices chosen. It is easily deduced from the arguments above that the only neighbours of  $z_j$  that are unseen by  $P_x$  are contained in  $V \cup Z \cup W$ . We know that  $N(x) \cap N(z) = V$ . Now choose  $u_{k+2} := v_i$  and  $u_{k+3} := z$ . If  $z$  is adjacent to  $z_j$ , pick  $u_{k+4} := z_j$ . Otherwise, pick  $u_{k+4} := w_1$  and  $u_{k+5} := z_j$ . In both cases, when  $z_j$  is chosen all of  $W \cup V$  have been seen. Therefore  $z_j$  has at most one neighbour (the vertex in  $Z \setminus \{z, z_j\}$ ) unseen by  $P_{z_j} \setminus \{z_j\}$ . This contradicts  $z$  being typical. Thus we have  $N(v_i) \cap N(w_j) = Z$ , completing the proof of the second statement of the lemma.  $\square$

We will now see that, for any  $v \in V(G_{max})$ , all but a bounded number of vertices in  $G_{max}$  are  $v$ -typical. We do this in the following manner. For each vertex  $v$  in  $G_{max}$  we will define a tree  $T(v)$  that will ‘explore’ the graph  $G_{max}$  outwards from  $v$ . As we will see, leaves on this tree will correspond to induced paths or cycles in  $G_{max}$  containing  $v$ . Every vertex on  $T$  will represent some vertex in  $G_{max}$  (and many vertices in  $T$  may represent the same vertex of  $G_{max}$ ). Our proof will proceed by showing that  $T$  must have a particular structure, which will in turn imply conditions on the structure of  $G_{max}$ .

As in Section 2, we define analogous concepts in an exploration tree  $T$  to some of those defined for a graph.

**Definition 3.6.** For  $F$  a finite graph and  $v \in V(F)$ , the *exploration tree from  $v$*  is a tree  $T = T(v)$  together with a function  $t : V(T) \rightarrow V(F)$  defined as follows.

- $T$  is a tree with vertex set  $V(T)$  disjoint from  $V(F)$ . The tree  $T$  is rooted at  $v_0 \in V(T)$  and we define  $t(v_0) := v$ .

- We define the tree in layers  $L_0, L_1, \dots$ , such that vertices in  $L_j$  will be at distance  $j$  from  $v_0$  in  $T$ . For any vertex  $v_i \in L_i$ , we let  $P_{v_i} := v_0, v_1, \dots, v_{i-1}, v_i$  be the unique path from  $v_0$  to  $v_i$  in  $T$ . The tree has the property that the set  $t(P_{v_i})$  will induce either a path or a cycle in  $F$ , where we write  $t(S) := \{t(x) : x \in S\}$  for any subset  $S \subseteq V(T)$  and  $t(H) := \{t(x) : x \in V(H)\}$  for any subgraph  $H \subseteq T$ .
- Given a set  $P \subseteq V(T)$  we say that it *sees* a vertex  $w \in V(F)$  if  $w \in N_F[t(P)]$ . If  $w \notin N_F[t(P)]$  we say  $w$  is *unseen* by  $P$ .
- Define  $L_0 := \{v_0\}$ . Given  $L_i$ , define  $L_{i+1}$  as follows. For each vertex  $x \in L_i$ :
  - If  $t(x) \in N_F(v)$  and  $i \neq 1$ , then  $x$  is a leaf of  $T$ ; it has no neighbours in  $L_{i+1}$ .
  - If  $t(x)$  has no neighbours in  $F$  that are unseen by  $V(P_x) \setminus \{x\}$  then  $x$  is a leaf of  $T$ .
  - Otherwise, let  $\{w_1, \dots, w_r\}$  be the set of vertices in  $N_F(t(x))$  unseen by  $P_x \setminus \{x\}$ . We define the neighbours of  $x$  in  $L_{i+1}$  to be a set of new vertices  $\{u_1, \dots, u_r\} \subseteq L_{i+1}$  and define  $t(u_i) := w_i$ . We call the vertices in  $N(x) \cap L_{i+1}$  *children* of  $x$ .

This process must terminate as  $F$  is finite.

Note that if some set  $P$  sees  $w$  then there exists  $u \in N_T[P]$  such that  $t(u) = w$ . Also observe that when we are constructing the tree, we allow neighbours of  $v_0$  to be children of vertices in later levels (and so disallow vertices seen by  $\{v_1, \dots, v_{i-1}\}$  rather than  $\{v_0, \dots, v_i\}$  when defining the children of  $x$  above). This allows a path terminating at a leaf  $l$  on the tree to correspond to an induced cycle containing  $v$ , since then  $t(l)$  is seen by  $v_0$  but not by  $v_i$  for  $i \geq 1$ . However, such a path will only correspond to an induced cycle containing  $v$  if  $t(l) \in N(v)$ .

As in Definition 2.6, we define a *descendant*, a *branch* and  $L(u)$  for  $u \in V(T)$  with respect to this tree. We will now describe a correspondence between certain leaves on  $T$  and induced cycles in  $F$ . Let  $L(T)$  denote the number of leaves on  $T$ . For  $z \in T$ , let  $L(z)$  be the number of leaves of  $T$  contained in  $B(z)$ .

**Lemma 3.7.** *For  $F$  a finite graph and  $v \in V(F)$ , let  $T$  be the exploration tree from  $v$ . We have  $f_v(F) \leq \frac{1}{2}L(T)$ .*

*Proof.* If  $f_v(F) = 0$ , the bound trivially holds. So suppose  $f_v(F) > 0$ . Let  $v, u_1, \dots, u_r, v$  be an induced cycle in  $F$ . By construction,  $T$  contains paths  $v_0, v_1, \dots, v_r$ , where  $t(v_i) = u_i$  and  $v_0, w_1, \dots, w_r$ , where  $t(w_i) = u_{r-i+1}$  (in both cases  $i \in \{1, \dots, r\}$ ). Both  $v_r$  and  $w_r$  are leaves of  $T$ . Thus an induced cycle containing  $v$  in  $F$  corresponds to exactly two paths in  $T$ . On the other hand, for a leaf  $v_r \in T$ , a path  $P_{v_r} = v_0, v_1, \dots, v_r$  in  $T$  corresponds to the set of vertices  $\{t(v_0), t(v_1), \dots, t(v_r)\} \subseteq V(F)$ . This set induces either a path or a cycle in  $F$ . Therefore each leaf in  $T$  can correspond to at most one induced cycle in  $F$  (there may be leaves that do not correspond to induced cycles). The result follows.  $\square$

We say that a vertex  $v \in T$  is *good* if it has exactly three children: call it *bad* otherwise. As we did for a graph, we define a game on  $T$  and use it to define vertices that are ‘(a)typical’ for  $T$ . The following definition is the analogue in  $T$  of Definition 3.4 for a graph.

**Definition 3.8.** Let  $F$  be a finite graph and let  $T$  be the exploration tree from  $v$  in  $F$ . Let  $w$  be a vertex in  $V(F) \setminus N^3[v]$ . We define the *w-typical-game* on  $T$  as follows. There are two players, Adversary and Builder. The game starts at vertex  $u_0 := v_0$  and the players choose a sequence of vertices  $\{u_1, u_2, \dots, u_k\} \subseteq V(T)$  under the following set of rules. At vertex  $u_i$ :

- If  $t(u_i) \in N^3[w]$ , then Adversary is the *active player*, otherwise Builder is.
- The active player chooses a child  $u_{i+1}$  of  $u_i$ .
- The game terminates when a vertex  $u_j$  is chosen such that  $u_j$  is a leaf.
- Adversary wins if either for some  $j$ , we have that  $t(u_j) \in N^3[w]$  and  $u_j$  is bad, or if upon termination of the game at vertex  $u_k$ , we have that there exists a vertex in  $N^3[w]$  that is unseen by  $\{u_0, \dots, u_k\}$ . Builder wins otherwise.

We say that a vertex  $w \in V(F) \setminus N^3[v]$  is *typical* for  $T$  if there exists a winning strategy for Builder in the *w-typical-game* on  $T$ . A vertex is *atypical* for  $T$  otherwise. It is clear that a vertex  $w$  being atypical for  $T$  means that Adversary has a strategy to ensure that, whatever strategy Builder chooses, either a bad vertex in  $N^3[w]$  must be chosen, or that there exists some vertex in  $N^3[w]$  that remains unseen by  $\{u_1, \dots, u_k\}$  upon termination at vertex  $u_k$ .

Now let  $c > 0$  and  $F$  be any  $n$ -vertex graph with  $f(F) \geq c \cdot 3^{n/3}$  and  $\Delta(F) \leq \Delta$ , for some constant  $\Delta$  ( $G_{max}$  satisfies these conditions as  $\Delta(G_{max})$  is bounded by Lemma 3.3). Our next aim is to prove that, for any vertex  $v \in V(F)$ , only a bounded number of vertices are atypical for  $T(v)$ . Using this fact with Lemma 3.5 will imply that the majority of the structure of  $G_{max}$  will be close to the structure of  $H_n$ . The remainder of the proof will consist of ‘cleaning’  $G_{max}$  to show that it must in fact be isomorphic to  $H_n$ .

We first outline how the proof will proceed before giving the details. We assume (in order to get a contradiction) that there is a large set  $A \subseteq V(F)$  of vertices atypical for  $T(v)$ , such that for each  $a, a' \in A$  we have  $N^3[a] \cap N^3[a'] = \emptyset$  (for any set of atypical vertices we can find a subset of constant proportion with this property as  $\Delta(F)$  is bounded). We will sequentially choose a path in  $T$  of vertices  $u_0, \dots, u_k$  where  $u_0 := v_0$  and  $u_k$  is a leaf.

For each  $a \in A$ , there exists a winning strategy  $\tau_a$  for Adversary in the *a-typical game* on  $T(v)$ . This means that whatever vertices  $v_i$  with  $t(v_i) \notin \bigcup_{a \in A} N^3[a]$ , are chosen in the path, for every  $a \in A$  we are able to ensure that either:

- (i) we choose a bad vertex  $v_i$  with  $t(v_i) \in N^3[a]$ , or
- (ii) there is some vertex in  $N^3[a]$  that remains unseen by  $\{u_0, \dots, u_k\}$  when the process terminates.

We will assume at the start that  $L(v_0)$  is bounded below by  $c3^{n/3}$ , for some constant  $c$ . As we move down the tree we keep track of the number of leaves that the branch we are in must contain. If we are at a vertex  $v_i$ , such that  $t(v_i)$  is not in  $N^3[a]$  for any  $a \in A$ , we choose the branch that has the most leaves. When  $t(v_i)$  is in  $N^3[a]$  for some  $a \in A$ , we play the winning strategy  $\tau_a$  to move towards the outcomes (i) or (ii), unless there is a sub-branch that contains a large proportion of the leaves in our current branch. These outcomes mean that the tree is ‘unbalanced’ in some way, and the strategy that achieves these outcomes picks branches that contain more leaves than average. As it turns out, when we reach a leaf and the process ends, if  $|A|$  was too large we find that the branch we are in ought to contain more than one leaf, a contradiction.

**Lemma 3.9.** *Let  $c > 0$  and  $\Delta > 0$  be fixed constants. Let  $\epsilon = 2^{-\Delta^{100}}$ . Let  $F$  be an  $n$ -vertex graph with  $\Delta(F) = \Delta$ . Let  $v \in V(F)$  and let  $T = T(v)$  be the exploration tree from  $v$  in  $F$  with root  $v_0$ . Let  $A \subseteq V(F) \setminus N^3[v]$  be a set of atypical vertices for  $T$  such that for all  $a, a' \in A$ , we have  $N^3[a] \cap N^3[a'] = \emptyset$ . If  $L(v_0) \geq c \cdot 3^{n/3}$ , then  $|A| < M$ , where  $M$  is the smallest integer such that  $c \cdot 3^{1/3}(1 + \epsilon)^M > \Delta$ .*

*Proof.* Suppose, in order to obtain a contradiction, that  $|A| \geq M$ . For each  $a \in A$ , Adversary has a winning strategy  $\tau_a$  played on vertices of  $N^3[a]$  in the  $a$ -typical game on  $T$ . As for all  $a, a' \in A$ , we have  $N^3[a] \cap N^3[a'] = \emptyset$ , these strategies are played on disjoint sets of vertices.

We sequentially choose a path  $v_0, u_1, \dots, u_k$  of vertices through the tree. We will pick  $u_1 \in N(v_0)$  such that  $L(u_1)$  is maximised and  $u_k$  will be a leaf. Define  $q_1 := \frac{c \cdot 3^{1/3}}{\Delta} 3^{(n-1)/3}$  and for each  $a \in A$  define  $C_1(a) := 1$ . At each subsequent stage, we will choose a vertex  $u_i$  and define quantities  $q_i$  and  $A_i$  (the subset of  $A$  that we still care about tracking). We will also define  $C_i(a)$  for each  $a \in A$ . We define

$$C_i := \frac{c \cdot 3^{1/3}}{\Delta} \prod_{a \in A} C_i(a) \text{ and } q_i := C_i 3^{\frac{n-m_i-1}{3}}, \quad (3.4)$$

where  $m_i$  is the number of vertices of  $V(F) \setminus \{v\}$  seen by  $V(P_{u_{i-1}}) \setminus \{u_0\}$  (thus  $n - m_i - 1$  vertices of  $V(F) \setminus \{v\}$  are unseen by  $V(P_{u_{i-1}}) \setminus \{u_0\}$ ). Throughout the process we will maintain the property that  $L(u_i) \geq q_i$  for each  $i$ .

We now describe an algorithm that determines our choice of vertices. For  $r \geq 1$ , let  $\epsilon_r = 2^{r-1}\epsilon$ .

---

#### Vertex Choice Algorithm

---

We pick  $u_1 \in N(v_0)$  such that  $L(u_1)$  is maximised, and define  $A_2 := A$ . Suppose the most recently chosen vertex is  $u_i$  and that  $m_i$  vertices of  $V(F) \setminus \{v\}$  have been seen by  $\{u_1, \dots, u_{i-1}\}$ . If  $v = u_i$  is not a leaf; we have two cases:

**Case 1:**  $t(v) \in N^3[a]$  for some  $a \in A_i$ .

Suppose it is the  $r$ -th time we have chosen a vertex  $y$  such that  $t(y) \in N^3[a]$ . We have two subcases:

**Subcase 1:**  $v$  is good.

In this case  $v$  has exactly three children  $y_1, y_2, y_3$ .

- (i) If there exists  $j$  such that  $L(y_j) \geq \frac{1}{3}(1 + \epsilon_r)C_i 3^{(n-m_i-1)/3}$  then choose  $u_{i+1} := y_j$ .
  - Set  $C_{i+1}(u) := (1 + \epsilon_r)C_i(u)$  and  $C_{i+1}(y) := C_i(y)$ , for all  $y \in A \setminus \{a\}$ .
  - Set  $A_{i+1} := A_i \setminus \{a\}$ .
- (ii) Else, every  $y_j$  must satisfy  $L(y_j) > \frac{1}{3}(1 - 2\epsilon_r)C_i 3^{(n-m_i-1)/3}$ . In this case, choose  $u_{i+1}$  according to strategy  $\tau_a$ .
  - Set  $C_{i+1}(a) := (1 - 2\epsilon_r)C_i(a)$  and  $C_{i+1}(y) := C_i(y)$ , for all  $y \in A \setminus \{a\}$ .
  - Set  $A_{i+1} := A_i$ .

**Subcase 2:**  $v$  is bad.

In this case  $v$  does not have exactly 3 children. Suppose  $v$  has children  $y_1, \dots, y_k$  for some  $k \neq 3$ . Pick  $j$  such that  $L(y_j)$  is maximised and set  $u_{i+1} := y_j$ .

- Set  $C_{i+1}(a) = (1 + \epsilon_r)C_i(a)$  and  $C_{i+1}(y) := C_i(y)$ , for all  $y \in A \setminus \{a\}$ .
- Set  $A_{i+1} := A_i \setminus \{u\}$ .

**Case 2:**  $t(v) \notin N^3[a]$  for all  $a \in A_i$ :

Then  $v$  has children  $y_1, \dots, y_k$  for some  $k \geq 1$ . Pick  $j$  such that  $L(y_j)$  is maximised and set  $u_{i+1} = y_j$ .

- Set  $C_{i+1}(y) := C_i(y)$ , for all  $y \in A$ .
- Set  $A_{i+1} := A_i$ .

The process terminates when  $u_i$  is a leaf.

We now analyse the consequences of choosing vertices in this manner.

**Claim 3.10.** For each vertex  $u_i$  chosen during the Vertex Choice Algorithm, we have  $L(u_i) \geq q_i$ .

*Proof of Claim 3.10.* We argue by induction on  $i$ : the case  $i = 1$  holds as we chose  $u_1 \in N(v_0)$  to maximise  $L(u_1)$ . Suppose  $L(u_i) \geq q_i = C_i 3^{(n-m_i-1)/3}$ . Now for the inductive step: we consider each case of the algorithm separately, and prove that the statement holds there.

In Subcase 1(i) we have:

$$L(u_{i+1}) \geq \frac{1}{3}(1 + \epsilon_r)C_i 3^{\frac{n-m_i-1}{3}} = C_{i+1} 3^{\frac{n-m_{i+1}-1}{3}} = q_{i+1}.$$

In Subcase 1(ii) we have:

$$L(y_j) > \frac{1}{3}(1 - 2\epsilon_r)C_i 3^{\frac{n-m_i-1}{3}} = C_{i+1} 3^{\frac{n-m_{i+1}-1}{3}} = q_{i+1}$$

In Subcase 2, recall that  $v$  has neighbours  $y_1, \dots, y_k$  (for  $k \neq 3$ ) and we pick  $u_{i+1}$  to be the  $y_j$  which maximises  $L(y_j)$ . Thus we have:

$$L(u_{i+1}) \geq \frac{C_i}{k} 3^{\frac{n-m_i}{3}} = C_i \frac{3^{k/3}}{k} 3^{\frac{n-m_i-k-1}{3}}.$$

The value of  $\frac{3^{k/3}}{k}$  is minimised for  $k \neq 3$  at  $k = 2$ . Thus,

$$L(u_{i+1}) \geq C_i \frac{3^{2/3}}{2} 3^{\frac{n-m_i-k-1}{3}} \geq C_i(1 + \epsilon_r) 3^{\frac{n-m_i-k-1}{3}} = C_{i+1} 3^{\frac{n-m_{i+1}-1}{3}} = q_{i+1}.$$

In Case 2, recall that  $v$  has neighbours  $y_1, \dots, y_k$  and we pick  $u_{i+1}$  to be the  $y_j$  which maximises  $L(y_j)$ . Thus we have:

$$L(u_{i+1}) \geq \frac{C_i}{k} 3^{\frac{n-m_i-1}{3}} = C_i \frac{3^{k/3}}{k} 3^{\frac{n-m_i-k-1}{3}} \geq C_i 3^{\frac{n-m_i-k-1}{3}} = C_{i+1} 3^{\frac{n-m_{i+1}-1}{3}} = q_{i+1},$$

where the last inequality is strict unless  $k = 3$ .  $\square$

We are now equipped to analyse what remains once the algorithm terminates at a leaf  $u_k$ . For each  $a \in A$  we have at least one of the following outcomes upon termination of the algorithm.

- (O1) During the algorithm, at a vertex  $a \in N^3[a]$ , we either chose a branch with a large proportion of leaves via Case 1(i) or we chose a bad vertex via Subcase 2.
- (O2) There is some vertex  $w \in N^3[a]$  that is unseen by  $V(P_{u_k})$  upon termination of the algorithm.

First observe that for all  $s \leq \Delta^3$ ,

$$(1 - 2\epsilon_1)(1 - 2\epsilon_2) \dots (1 - 2\epsilon_s)(1 + \epsilon_{s+1}) > (1 + \epsilon). \quad (3.5)$$

Our choice of  $\epsilon$  ensures that each factor on the left hand side is greater than zero.

Suppose  $a \notin A_k$ . Then there exists some  $j$  such that  $a \in A_j$  but  $a \notin A_{j+1}$ . Thus at the  $j$ th stage of the algorithm we had  $t(u_j) \in N^3[a]$  and we either chose a branch with a large proportion of leaves via Case 1(i) or we chose a bad vertex via subcase 2. Let  $t := |t(\{u_1, \dots, u_j\}) \cap N^3[a]|$  be the number of vertices,  $x \in T$  such that  $t(x) \in N^3[a]$ , that we have chosen up to the  $j$ th stage. It is clear from the algorithm that

$$C_k(a) = C_{j+1}(a) = (1 - 2\epsilon_1)(1 - 2\epsilon_2) \dots (1 - 2\epsilon_t)(1 + \epsilon_{t+1}). \quad (3.6)$$

We have  $t + 1 \leq \Delta^3$  (as  $|N^3[a]| \leq \Delta^3$  by hypothesis), and so by (3.5) we have in this case that

$$C_k(a) > (1 + \epsilon). \quad (3.7)$$

Now suppose  $a \in A_k$ . We pass through Case 1 at most  $\Delta^3$  times, as by hypothesis  $|N^3[a]| \leq \Delta^3$ . Thus, for each such  $a$ , we have  $C_k(a) \geq \prod_{i=1}^{\Delta^3} (1 - 2\epsilon_i)$ . By choice of  $\epsilon$  we have that for all  $s \leq \Delta^3$ ,

$$3^{1/3} > 1 + 2^{2(s+1)}\epsilon,$$

so by (3.5),

$$3^{1/3} \cdot C_k(a) > (1 + \epsilon). \quad (3.8)$$

For each  $a \in A_k$ , the set  $V(P_{u_k})$  does not see all of  $N^3[a]$ , as we must either have outcome (O1) or (O2) for  $a$ , and if we achieved (O1), then  $a$  would not be in  $A_k$ . So at termination we have  $n - m_k - 1 \geq |A_k|$  and so by the definition of  $q_k$  (3.4) we have:

$$q_k \geq C_k 3^{\frac{|A_k|}{3}}. \quad (3.9)$$

By (3.7), we have:

$$\prod_{a \in A} C_k(a) \geq (1 + \epsilon)^{|A \setminus A_k|} \prod_{a \in A_k} C_k(a),$$

and so by substituting for  $C_k$  in (3.9) and applying (3.8), we have:

$$q_k \geq \frac{c \cdot 3^{1/3}}{\Delta} 3^{\frac{|A_k|}{3}} \prod_{a \in A} C_k(a) \geq \frac{c \cdot 3^{1/3}}{\Delta} (1 + \epsilon)^{M - |A_k|} 3^{\frac{|A_k|}{3}} \prod_{a \in A_k} C_k(a) \geq \frac{c \cdot 3^{1/3}}{\Delta} (1 + \epsilon)^M > 1.$$

This contradicts Claim 3.10, as  $u_k$  is a leaf and thus  $L(u_k) = 1$ . This concludes the proof of Lemma 3.9. □

**Corollary 3.11.** *Let  $c > 0$  and  $\Delta > 0$  be fixed constants. Let  $F$  be an  $n$ -vertex graph with  $f(F) \geq 2c \cdot 3^{n/3}$  and  $\Delta(F) = \Delta$ . Then there exists some  $v \in V(F)$  such that  $O(1)$  vertices are atypical for  $T(v)$ , the exploration tree from  $v$  in  $F$ .*

*Proof.* Let  $A$  be the set of vertices that are atypical for  $T$ . Let  $U$  be the largest subset of  $A$  such that for all  $a, a' \in U$ , we have  $N^3[a] \cap N^3[a'] = \emptyset$ . As  $|N^3[x]| \leq \Delta^3$  for every  $x \in V(F)$ , we have

$$|U| \geq \frac{|A|}{\Delta^3}. \quad (3.10)$$

We wish to apply Lemma 3.9 to  $F$ ,  $T$  and  $U$ .

As  $f(F) \geq 2c \cdot m(n)$ , by Lemma 3.2 there exists  $v \in V(F)$  such that  $f_v(F) \geq \frac{c}{10} m(n)$  for all  $v \in F$ . Combining this with Lemma 3.7 gives

$$L(v_0) = L(T) \geq \frac{1}{2} f_v(F) \geq \frac{c}{10} m(n) = c' 3^{n/3},$$

for some constant  $c'$  and where the last inequality follows from 3.3.

So we can apply Lemma 3.9 to get that  $|U| < M$ , where  $M$  is the smallest integer such that  $c \cdot 3^{1/3} (1 + 2^{-\Delta^{100}})^M > \Delta$ . Combining this with (3.10) gives  $|A| = O(1)$ . □



In particular, using Lemma 3.2, the proof works for every  $v \in V(G_{max})$ . So for any  $v \in V(G_{max})$ ,  $O(1)$  vertices are atypical for  $T(v)$ .

Let  $\mathcal{B} := (B_1, \dots, B_k)$  be a braid in  $G_{max}$ . If  $|B_i| = 3$  for all  $i$ , we call  $\mathcal{B}$  a *3-braid*. For a braid  $\mathcal{B}$  of length at least 4, we say that an induced cycle *passes through*  $\mathcal{B}$  if it contains a vertex from every cluster of  $\mathcal{B}$ . Call a braid *maximal* if it is not contained in any longer braid. We require the following simple deduction.

**Lemma 3.12.**  *$G_{max}$  contains  $O(1)$  maximal 3-braids and a 3-braid  $\mathcal{B}$  such that  $|V(\mathcal{B})| = \Omega(n)$ . Moreover, for any 3-braid  $\mathcal{B}'$  on  $rn$  vertices, at least  $f(H_n) (1 - 3^{-rn/6})$  induced cycles in  $G_{max}$  pass through  $\mathcal{B}'$ .*

*Proof.* Let  $v \in V(G_{max})$ . The only vertices which can be contained in more than one maximal 3-braid lie in end clusters. By Lemma 3.5, every  $v$ -typical vertex is contained in a central cluster of exactly one maximal 3-braid. So any vertex in the end cluster of a maximal 3-braid must be  $v$ -atypical. By Corollary 3.11,  $O(1)$  vertices are  $v$ -atypical for  $G_{max}$ . Each of these vertices is contained in at most  $\Delta(G_{max}) \leq 30$  maximal 3-braids. Thus  $G_{max}$  contains  $O(1)$  maximal 3-braids.

The union of the maximal 3-braids contains all the typical vertices and so it must contain at least  $n - O(1)$  vertices, which is at least  $n/2$  for large  $n$ . Therefore, when  $n_0$  is sufficiently large, some 3-braid  $\mathcal{B} = (B_1, \dots, B_k)$  must contain  $\Omega(n)$  vertices.

Now, for the second claim, observe that if a cycle does not pass through  $\mathcal{B}$  then it is either a  $C_4$  contained in  $\mathcal{B}$  (there are at most  $O(n^4)$  of these), or it is contained in  $V(G_{max}) \setminus \bigcup_{i=3}^{k-2} B_i$  (by Lemma 3.1, there are at most  $[(1-r)n + 12] \binom{30^3}{2} 3^{\lfloor (1-r)n+9 \rfloor / 3}$  of these). Therefore at most

$$[(1-r)n + 12] \binom{30^3}{2} 3^{\lfloor (1-r)n+9 \rfloor / 3} + O(n^4)$$

induced cycles of  $G_{max}$  do not pass through  $\mathcal{B}$ . So for  $n_0$  sufficiently large, at least

$$f(H_n) (1 - 3^{-rn/6})$$

induced cycles pass through  $\mathcal{B}$ . □

The next lemma determines almost entirely the structure of  $G_{max}$ .

**Lemma 3.13.**  *$G_{max}$  is a cyclic braid.*

*Proof.* Suppose  $G_{max}$  is not a cyclic braid, and choose  $v$  to be an atypical vertex of  $G_{max}$ . Let  $\mathcal{B} := (B_1, \dots, B_{Cn/3})$  be the longest 3-braid in  $G_{max}$ . Let  $Q$  be the number of induced cycles in  $G_{max}$  that pass through  $\mathcal{B}$ . Again using Lemma 3.12, we know

$$Q \geq f(H_n) (1 - 3^{-Cn/6}). \quad (3.11)$$

Now let  $G' := G[V(G_{max}) \setminus \bigcup_{i=2}^{Cn/3-1} B_i]$ . Let  $x$  and  $y$  be two new vertices and define  $H$  to be the graph on vertex set  $V(H) := V(G') \cup \{x, y\}$ , and edge set

$$E(H) := E(G') \cup \{xb : b \in B_1\} \cup \{yb : b \in B_{Cn/3}\}.$$

We have

$$Q = 3^{(Cn-6)/3} p_2(H; x, y). \quad (3.12)$$

Combining (3.11) and (3.12) gives

$$p_2(H; x, y) \geq 3^{-(Cn-6)/3} \cdot f(H_n) (1 - 3^{-Cn/6}). \quad (3.13)$$

We now focus on the structure of  $H$ . Let us call a central cluster  $C$  of a maximal 3-braid in a graph *supercentral* if for any  $x \in C$  and  $y$  in an end cluster,  $d(x, y) \geq 3$ . We define a new graph  $H'$  via the following process.

- Set  $F_1 := H$ .
- Let  $i$  be maximal such that we have defined  $F_i$ . Suppose there exists a vertex  $v_i \in F_i$ , contained in a supercentral cluster  $C_i$  of a maximal 3-braid  $\mathcal{B}_i$ . Suppose  $C_i$  is adjacent to clusters  $C_1^i$  and  $C_2^i$ . We define  $F_{i+1}$  to be the graph obtained from  $F_i$  by deleting  $C_i$  and adding every edge  $\{uw : u \in C_1^i, w \in C_2^i\}$ .
- If there exists no such vertex  $v_i$ , we define  $H' := F_i$ .

The process will terminate as  $H$  has a finite number of vertices. Observe that  $F_{i+1}$  is a braid if and only if  $F_i$  is a braid. In addition, when  $H'$  is a braid it is easy to reverse this process to find  $H$ . We will show that  $H'$  is a braid.

Any  $v$ -typical vertex in  $F_i$  that does not get deleted during the process is  $v$ -typical in  $F_{i+1}$ . Hence any  $v$ -typical vertex in  $G_{max}$  that does not get deleted is  $v$ -typical in  $H'$ . By Lemma 3.12,  $G_{max}$  contains  $O(1)$  maximal 3-braids. At most 18 vertices from each of these braids will remain in  $H'$  when the process terminates. Any vertex not contained in a 3-braid in  $G_{max}$  is  $v$ -atypical in  $G_{max}$ . By Corollary 3.11 there are  $O(1)$  such vertices. So as  $H'$  contains all the atypical vertices of  $G_{max}$  and at most 18 vertices from each 3-braid in  $G_{max}$ , we have  $|V(H')| = O(1)$ .

At stage  $i$  of the process,  $F_{i+1}$  contains all induced cycles of  $F_i$  that do not pass through  $\mathcal{B}_i$  and a third of the number of cycles in  $F_i$  that do pass through  $\mathcal{B}_i$ . Thus we have

$$p_2(F_{i+1}; x, y) \geq 3^{-1} \cdot p_2(F_i; x, y), \quad (3.14)$$

and so

$$p_2(H'; x, y) \geq 3^{-(|V(H)|-|V(H')|)/3} \cdot p_2(H; x, y), \quad (3.15)$$

Combining (3.13) and (3.15) and observing that  $|V(H)| = (1 - C)n + 8$  gives

$$p_2(H'; x, y) \geq 3^{(-n-2+|V(H')|)/3} \cdot f(H_n) (1 - 3^{-Cn/6}). \quad (3.16)$$

As

$$3^{(-n-2+|V(H')|)/3} f(H_n) = f_2(|V(H')|),$$

when  $n_0$  is sufficiently large we have

$$3^{(-n-2+|V(H')|)/3} \cdot f(H_n) (1 - 3^{-Cn/6}) > f_2(|V(H')|) - 1. \quad (3.17)$$

As  $p_2(H'; x, y)$  is an integer, by taking  $n_0$  to be sufficiently large, (3.16) and (3.17) give

$$p_2(H'; x, y) \geq f_2(|V(H')|).$$

Therefore, by Theorem 2.1,  $H'$  must be isomorphic to a graph in  $\mathcal{F}(|V(H')|)$ . Thus  $H'$  is a braid. We can reverse the process we applied above (adding back in the supercentral clusters) to recreate  $H$  from  $H'$ . We see that  $H$  is a graph in  $\mathcal{F}(|V(H)|)$ , and hence  $G_{max}$  is a cyclic braid.  $\square$

**Corollary 3.14.** *We have the following:*

- *when  $n \equiv 0$  modulo 3,  $G_{max}$  has exactly  $n/3$  clusters of size 3;*
- *when  $n \equiv 1$  modulo 3,  $G_{max}$  has either one cluster of size 4 and  $(n-4)/3$  of size 3, or two of size two and  $(n-4)/3$  of size 3;*
- *when  $n \equiv 2$  modulo 3,  $G_{max}$  has exactly one cluster of size 2 and  $(n-2)/3$  of size 3.*

We are now in a position to complete the proof of Theorem 1.2. In our next lemma, we will show that the clusters in  $G_{max}$  cannot contain any edges, and thus we will prove the required result for  $n \equiv 0$  or 2 modulo 3. In the remaining case we will need a short argument to decide whether the graph contains two clusters of size two, or one of size 4. In both cases, the arguments are essentially routine checks.

**Lemma 3.15.** *No cluster of  $G_{max}$  contains any edges.*

*Proof.* First observe, that if  $e$  is an edge within a cluster, the only induced cycles containing  $e$  can be triangles, either contained within the cluster, or containing exactly one vertex from a neighbouring cluster; or induced copies of  $C_4$  within the cluster (in the case that the cluster contains 4 vertices).

Let  $B$  be a cluster adjacent to clusters  $A$  and  $C$ . Suppose there exists an edge  $e = uv$  where  $u, v \in V(B)$ . The edge  $e$  is contained in at most  $|A| + |C| + (|B| - 2)$  induced cycles within  $G_{max}$ . The graph  $G' = G_{max} \setminus \{e\}$  will contain at least  $|A||C|$  induced copies of  $C_4$  (for any  $x \in A, y \in C$ , the set  $\{x, y, u, v\}$  induces a  $C_4$ ) that are not induced cycles in  $G_{max}$ .

As  $G_{max}$  cannot contain both a cluster of size 2 and a cluster of size 4, we have

$$|A||C| > |A| + |C| + (|B| - 2),$$

unless  $|B| = 3$  and at least one of  $|A|$  or  $|C|$  is equal to 2. Except for this case, the number of induced cycles in  $G' = G_{max} \setminus \{e\}$  is greater than the number of induced cycles in  $G_{max}$ , a contradiction.

Now assume  $|B| = 3$  and suppose without loss of generality that  $A = \{a_1, a_2\}$ . First consider the case where  $|C| = 3$ . Suppose  $B$  contains an edge  $e = uv$ . This edge is contained in at most 6 triangles in  $G_{max}$ . By the above argument,  $A$  does not contain an edge. The graph  $G' = G_{max} \setminus \{e\}$  will contain at least 7 induced copies of  $C_4$  that are not induced cycles

in  $G_{max}$  (for any  $x \in A$ ,  $y \in C$ , the sets  $\{x, y, u, v\}$  and  $\{a_1, a_2, u, v\}$  induce copies of  $C_4$ ). Thus  $f(G') > f(G_{max})$ , a contradiction.

Now assume that  $C = \{c_1, c_2\}$ . If  $B$  contains an edge  $e$ , this edge is contained in at most 5 triangles in  $G_{max}$ . The graph  $G' = G_{max} \setminus \{e\}$  contains at least 6 induced copies of  $C_4$  that are not induced cycles in  $G_{max}$ . Thus  $f(G') > f(G_{max})$ , a contradiction. So no cluster in  $G_{max}$  contains an edge.  $\square$

We have proved Theorem 1.2 in the cases where  $n \equiv 0$  or  $2$  modulo  $3$ . It remains to prove the result in the case  $n \equiv 1$  modulo  $3$ .

**Lemma 3.16.** *When  $n \equiv 1$  modulo  $3$ ,  $G_{max}$  is isomorphic to  $H_n$ .*

*Proof.* By Corollary 3.14 and Lemma 3.15, we know that  $G_{max}$  is one of two empty cyclic braids. One possibility is that it is isomorphic to  $H_n$ . The other possibility is that  $G_{max}$  is an empty cyclic braid  $G_2$  with exactly two clusters of size 2, and the rest of size 3. An induced cycle in  $H_n$  or  $G_2$  either contains exactly one vertex from each cluster, or is an induced copy of  $C_4$ . In both  $H_n$  and  $G_2$ , the number of cycles containing exactly one vertex from each cluster is  $4 \cdot 3^{(n-4)/3}$ . Thus the only difference in the number of induced cycles comes from the number of induced copies of  $C_4$ .

There are two types of  $C_4$ . Type 1 contains vertices from exactly two clusters. Type 2 contains vertices from three clusters.  $H_n$  contains  $3(n+5)$  induced type 1 cycles;  $G_2$  contains at most  $3n-14$  of this form (fewer if the two clusters of size 2 are not adjacent). The graph  $H_n$  contains  $9(n+4)$  induced type 2 cycles;  $G_2$  contains at most  $9n-42$  of this form. Thus it is clear that  $H_n$  contains more induced cycles than  $G_2$  and therefore  $G_{max}$  must be isomorphic to  $H_n$ .  $\square$

## 4 Proof of Theorem 1.5

The proof of Theorem 1.5 follows the same lines as the proof of Theorem 1.2. Before proceeding with the details of the proof, we first give an outline of what is to come. Let  $0 < \alpha < 1$  be any constant and let  $F$  be an  $n$ -vertex graph containing at least  $\alpha \cdot m(n)$  induced cycles. We will show that we can delete  $O(1)$  vertices from  $F$  to give a graph  $F'$  such that  $\Delta(F) = O(1)$ . We can then apply Lemma 3.9 to  $F'$  to show that  $F'$  contains a bounded number of atypical vertices. The result will immediately follow. We cannot simply apply Lemma 3.9 to  $F$ , as  $F$  may contain vertices of arbitrarily large degree.

**Lemma 4.1.** *Let  $0 < \alpha < 1$  and define  $\Delta^* = \Delta^*(\alpha)$  to be the smallest integer such that  $3\binom{\Delta^*}{2}3^{(1-\Delta^*)/3} < \alpha \cdot 4 \cdot 3^{-4/3}$ . Let  $F$  be an  $n$ -vertex graph with  $f(F) \geq \alpha \cdot m(n)$ . Then we can delete  $O(1)$  vertices from  $F$  to give a graph  $H$  with  $\Delta(H) \leq \Delta^*$ . Moreover,  $f(H) \geq \frac{\alpha}{2} \cdot m(n)$ .*

*Proof.* Suppose that  $\Delta(F) > \Delta^*$  (else we are trivially done). We create a new graph  $H$ , with  $\Delta(H) \leq \Delta^*$ , in the following manner. Define  $F_1 := F$ . Let  $i$  be maximal such that  $F_i$  has been defined. If there exists a vertex  $v_i \in V(F_i)$  with  $d(v_i) > \Delta^*$  then define

$F_{i+1} := F_i \setminus \{v_i\}$ . This process will terminate as  $F$  has a finite number of vertices. Suppose the process terminates at a graph  $F_j$  satisfying  $\Delta(F_j) \leq \Delta^*$ . Define  $H := F_j$ .

To prove the first statement of the lemma it suffices to show that  $j = O(1)$ . In this case, we could create a graph  $H$  with  $\Delta(H) \leq \Delta^*$  by deleting all edges incident to  $\{v_1, \dots, v_j\}$ . To prove that  $j = O(1)$ , we will bound the size of  $f(F) - f(H)$  and use this to show that unless  $j = O(1)$ , we have  $f(F) < \alpha \cdot m(n)$ , a contradiction.

By Lemma 3.1,

$$f_{v_i}(F_i) \leq \binom{d_{F_i}(v_i)}{2} 3^{(n-i-d_{F_i}(v_i)-1)/3} \leq \binom{\Delta^*}{2} 3^{(n-i-\Delta^*-1)/3},$$

where the second inequality follows as the function  $\binom{x}{2} 3^{-x/3}$  is decreasing for  $x \geq 6$ . So

$$\begin{aligned} f(H) &= f(F) - \sum_{i=1}^{j-1} f_{v_i}(F_i) \\ &\geq f(F) - \binom{\Delta^*}{2} 3^{(n-\Delta^*+1)/3} \sum_{i=1}^{j-1} 3^{-i/3} \\ &\geq f(F) - 3 \binom{\Delta^*}{2} 3^{(n-\Delta^*+1)/3} \end{aligned} \tag{4.1}$$

As  $|V(H)| = n - j + 1$ , by (3.3) we have  $f(H) = c \cdot 3^{(n-j+1)/3}$  for some constant  $c$ . Combining this with (4.1) gives:

$$f(F) \leq c \cdot 3^{(n-j+1)/3} + 3 \binom{\Delta^*}{2} 3^{(n-\Delta^*+1)/3}. \tag{4.2}$$

There exists a constant  $J$  such that, whenever  $j \geq J$ , for  $n$  sufficiently large we have

$$c \cdot 3^{(n-j+1)/3} < \frac{1}{2} (\alpha \cdot 4 \cdot 3^{-4/3}) 3^{n/3}. \tag{4.3}$$

Suppose that  $j \geq J$  and let  $n$  be sufficiently large. Using the definition of  $\Delta^*$  and substituting (4.3) into (4.2), gives

$$f(F) < \alpha \cdot 4 \cdot 3^{-4/3} 3^{n/3} < \alpha \cdot m(n),$$

where the final inequality is implied by (3.1). This contradicts the hypothesis that  $f(F) \geq \alpha \cdot m(n)$ . Therefore  $j < J$  and in particular,  $j = O(1)$ , completing the proof of the first statement of the lemma.

We now prove the second statement. By (4.1) we have

$$f(H) \geq f(F) - \frac{1}{2} \binom{\Delta^*}{2} 3^{(n-\Delta^*+1)/3}.$$

Given the definition of  $\Delta^*$ , it is easy to see that  $f(H) \geq \frac{\alpha}{2} m(n)$ . □

*Proof of Theorem 1.5.* Let  $F$  be an  $n$ -vertex graph containing at least  $\alpha \cdot m(n)$  induced cycles. By Lemma 4.1, we can delete  $O(1)$  vertices from  $F$  to give a graph  $F'$  with  $\Delta(F') \leq \Delta^* = \Delta^*(\alpha)$  and  $f(F') \geq \frac{\alpha}{2}m(n)$ . Now pick  $v \in V(F')$  such that  $f_v(F) \geq \frac{\alpha}{20}$  (possible by Lemma 3.2) and let  $A$  be the set of  $v$ -atypical vertices in  $F'$ . By applying Lemma 3.9 we deduce that  $|A| = O(1)$  (where  $M$  is defined as in Lemma 3.9). By Lemma 3.5, every  $v$ -typical vertex in  $F'$  is contained in a central cluster of exactly one maximal 3-braid. Clearly we can obtain  $H$  from  $F'$  by adding and deleting edges incident to vertices in  $A$ . The result follows.  $\square$

## 5 Induced odd or even cycles

In this section we will prove Theorem 1.7. The proofs of Theorem 1.8 and Theorem 1.9 will closely follow.

Given a graph  $G$ , define  $f_o(G)$  to be the number of induced odd cycles contained within  $G$ . Similarly, for  $v \in G$ , define  $f_o^v(G)$  to be the number of induced odd cycles in  $G$  that contain  $v$ . It is clear that:

$$m_o(n) \geq f_o(G_n) = \Omega(3^{n/3}), \quad (5.1)$$

where  $G_n$  is defined as in Section 1. The proof of Theorem 1.7 will follow from Theorem 1.5 and some arguments analogous to those used in Theorem 1.2. For the latter, we refer back to Sections 2 and 3 where necessary. The main difference is that, instead of applying Theorem 2.1, we use Theorem 2.9.

We fix a large constant  $n_0$  and let  $G$  be a graph on  $n \geq n_0$  vertices that contains  $m_o(n)$  induced odd cycles. In what follows we will let  $n_0$  be sufficiently large when required and we will make no attempts to optimise the constants given in our argument.

*Sketch proof of Theorem 1.7.* We first show that  $\Delta(G) = O(1)$  using analogous arguments to those in Theorem 1.2.

Lemma 3.2 holds (as (5.1) gives us the analogous bound to (3.1) that we need). Thus every vertex is contained in at least  $\frac{1}{20}m_o(n)$  induced odd cycles. Thus we have

$$m_o(n+1) \geq \left(1 + \frac{1}{20}\right) m_o(n), \quad (5.2)$$

as in (3.2).

We use the same argument as in Lemma 3.3, replacing  $m(n)$  with  $m_o(n)$ , to show that  $\Delta(G) \leq 35$  (we get a different value for  $\Delta$  as we use the lower bound  $m_o(G_n) \geq 3^{(n-8)/3}$  and this differs from the lower bound we used for  $m(n)$ ).

By (5.1) we have  $f_o(G_n) = \Omega(3^{n/3})$ . So we can apply Theorem 1.5 to show we can add and delete edges incident to  $O(1)$  vertices of  $G$  to give a cyclic braid  $H$  with the same cluster sizes as  $H_n$ . Using this and the knowledge that  $\Delta(G) = O(1)$ , it is easily seen that  $G$  must contain a 3-braid  $\mathcal{B}$  of even length such that  $|V(\mathcal{B})| = rn$  for some constant  $r$ .

We now show that  $G$  must be a cyclic braid. We can use the same argument as in Lemma 3.13. We now follow essentially the same proof as in Theorem 1.2 with  $f_o(G_n)$  in place of

$f(H_n)$ . However, when we apply the process of deleting central clusters, we delete a pair of adjacent clusters at a time (to maintain the count of odd cycles). We again reach a graph  $H'$  with  $|V(H')| = O(1)$ . We can make the analogous deductions from there to reach the bound

$$f_2^o(n) \leq p_2^o(H'; x, y).$$

We then apply Theorem 2.9 to determine that  $H' \in \mathcal{F}_o(|V(H')|)$ . Reversing the process of deleting central clusters to obtain  $H$  from  $H'$ , we see that  $H \in \mathcal{F}_o(|V(H)|)$ . Therefore  $G$  is a cyclic braid, with clusters all of size 3 except:

- three clusters of size 2, when  $n \equiv 0$  modulo 6;
- two clusters of size 2, when  $n \equiv 1$  modulo 6;
- one cluster of size 2, when  $n \equiv 2$  modulo 6;
- a single cluster of size 4, when  $n \equiv 4$  modulo 6; and
- either two clusters of size 4 or four clusters of size 2, when  $n \equiv 5$  modulo 6

It remains to determine whether there are edges within the clusters, the relative positions of the clusters in the cyclic braid (in the cases where more than one cluster does not have size 3) and, in the case  $n \equiv 5$  modulo 6, to determine the precise cluster sizes. Using arguments of a similar nature to those in Lemma 3.15 and Lemma 3.16, it can be easily checked that  $G \cong G_n$  for every value of  $n \geq n_0$ .  $\square$

Theorem 1.7 determines which  $n$ -vertex graphs contain the maximum number of odd cycles. Following essentially the same argument we can prove Theorem 1.8 and Theorem 1.9, which determine the family of  $n$ -vertex graphs that contain the maximum number of odd holes or even holes respectively.

*Proof of Theorem 1.8 and Theorem 1.9.* We use the same argument as in the proof of Theorem 1.7. In the case of Theorem 1.9, it is easy to modify the argument to consider odd induced cycles rather than even. The main difference is at the final stage, where we know  $G$  is a cyclic braid and the possible cluster sizes in  $G$ . Changing the positions of clusters and edges within clusters can only affect the holes that do not contain a vertex from every cluster. Thus any hole that can be affected must have size 3 or 4.

For the odd case, the positions of the clusters and the existence of edges within clusters will not alter the number of odd holes (as any induced cycle with size 3 or 4 is not an odd hole). In the even case, it is easy to check that  $G$  must be isomorphic to  $E_n$  to maximise the number of holes of size 4 given the cluster sizes.  $\square$

## 6 Conclusion

For sufficiently large  $n$ , we have determined precisely which graphs on  $n$  vertices contain the maximum number of induced cycles, the maximum number of odd or even induced cycles, and the maximum number of holes. However, there are a number of interesting related questions.

In our proofs above we make no attempts to optimise the value of  $n_0$ . We know that in some small cases,  $H_n$  does not contain the maximum number of induced cycles [10, 13]. We believe Theorem 1.2 ought to be true for  $n_0 = 30$ , but our proof gives a much larger number. There are several places where we could improve the bound, most notably by choosing a more careful strategy in Lemma 3.9. However we omit the details as the bound would still be extremely large.

It is natural to consider induced cycles of some length that depends on  $n$ . Let  $c(n, l)$  be the maximum number of length  $l$  induced cycles that can be contained in a graph on  $n$  vertices. Let  $C(n, l)$  be the set of graphs containing  $c(n, l)$  induced cycles of length  $l$ .

**Question 6.1.** For  $l = l(n)$ , what is  $C(n, l)$ ?

When  $l$  is linear we believe the following should hold.

**Conjecture 6.2.** Fix  $c \in (0, 1)$ . If  $l(n) = \lceil cn \rceil$ , then for sufficiently large  $n$  the only graphs in  $C(n, l)$  are cyclic braids of length  $l$ .

We hope to return to this in a later paper. Perhaps a similar result holds down to cycles of length  $\Omega(\sqrt{n})$ .

**Question 6.3.** Suppose  $l(n) > \sqrt{n}$ . For sufficiently large  $n$ , are all graphs in  $C(n, l)$  cyclic braids?

Another related question is to ask about induced subgraphs which are subdivisions of some fixed graph  $H$ .

**Question 6.4.** Given a fixed finite graph  $H$ , what is the maximum number of induced subdivisions of  $H$  that can be contained in a graph on  $n$  vertices (and which graphs realise this maximum)?

Theorem 1.2 answers this question for  $H = C_3$ , but what happens for other graphs? For instance, which graphs maximise the number of induced subdivisions of  $K_{1,3}$ ? For large  $n$ , are the extremal graphs always blowups of some subdivision of  $H$ ? The rooted version of the question is also interesting, where we consider induced subdivisions of  $H$  where the branch vertices are fixed (for instance Theorem 2.1 is a result of this form for  $H = K_2$ ).

Finally, we remark that the related problem of finding the graph on  $n$  vertices that contains the most cycles (not necessarily induced) is trivial as we can just take  $K_n$ . However, the problem becomes interesting when we forbid certain subgraphs. Arman, Gunderson and Tsaturian [1] proved that, for sufficiently large  $n$ , the  $n$ -vertex triangle free graph containing the most cycles is a balanced complete bipartite graph. More recently, Morrison, Roberts



and Scott [11] extended this to a wider range of forbidden subgraphs: For  $k \geq 2$ ,  $H$  an edge-critical graph with chromatic number  $\chi(H) = k + 1$ , and  $n$  sufficiently large, the unique  $n$ -vertex  $H$ -free graph which contains the most cycles is the complete balanced  $k$ -partite graph.

## 7 Acknowledgements.

We would like to thank Brendan McKay and Mike Robson for helpful discussions. We would also like to thank Andrew Treglown for bringing [14] to our attention.

## References

- [1] A. Arman, D. Gunderson, and S. Tsaturian, *Triangle-free graphs with the maximum number of cycles*, arXiv:1501.01088v2, preprint, March 2015.
- [2] J. Balogh, P. Hu, B. Lidický, and F. Pfender, *Maximum density of an induced 5-cycle is achieved by an iterated blow-up of a 5-cycle*, arXiv:1411.4645v1, preprint, November 2014.
- [3] Béla Bollobás, Yoshimi Egawa, Andrew Harris, and Guo Ping Jin, *The maximal number of induced  $r$ -partite subgraphs*, Graphs Combin. **11** (1995), no. 1, 1–19.
- [4] Béla Bollobás, Chiê Nara, and Shun-ichi Tachibana, *The maximal number of induced complete bipartite graphs*, Discrete Math. **62** (1986), no. 3, 271–275.
- [5] Jason I. Brown and Alexander Sidorenko, *The inducibility of complete bipartite graphs*, J. Graph Theory **18** (1994), no. 6, 629–645.
- [6] Chaim Even-Zohar and Nati Linial, *A note on the inducibility of 4-vertex graphs*, Graphs Combin. **31** (2015), no. 5, 1367–1380.
- [7] Geoffrey Exoo, *Dense packings of induced subgraphs*, Ars Combin. **22** (1986), 5–10.
- [8] Hamed Hatami, James Hirst, and Serguei Norine, *The inducibility of blow-up graphs*, J. Combin. Theory Ser. B **109** (2014), 196–212.
- [9] James Hirst, *The inducibility of graphs on four vertices*, J. Graph Theory **75** (2014), no. 3, 231–243.
- [10] Brendan McKay, personal communication.
- [11] N. Morrison, A. Roberts, and A. Scott, *Maximising the number of cycles in  $H$ -free graphs*, In preparation.

- [12] Nicholas Pippenger and Martin Charles Golumbic, *The inducibility of graphs*, J. Combinatorial Theory Ser. B **19** (1975), no. 3, 189–203.
- [13] Mike Robson, personal communication.
- [14] Z. Tuza, *Unsolved combinatorial problems.*, BRICS Lecture Series, LS-01-1 (2001), (available from <http://www.brics.dk/publications/>).